Structural and functional characterization of protein complexes in the blood coagulation cascade

Madsen, Jesper Jonasson

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Structural and functional characterization of protein complexes in the blood coagulation cascade

Jesper Jonasson Madsen

August 2014

Danmarks Tekniske Universitet

Structural and functional characterization of protein complexes in the blood coagulation cascade

Rapport version: August 2014

Af Jesper Jonasson Madsen

Udgivet af: Ledelse og Administration, Anker Engelunds Vej 1, Bygning 101 A, 2800 Kgs. Lyngby

Rekvireres: www.dtu.dk

ISSN: [0000-0000] (elektronisk udgave)

ISBN: [000-00-0000-000-0] (elektronisk udgave)

ISSN: [0000-0000] (trykt udgave)

ISBN: [000-00-0000-000-0] (trykt udgave)

-

TECHNICAL UNIVERSITY OF DENMARK
Department of Chemistry
Physical and Biophysical Chemistry

Dissertation Examination Committee:
Kasper P. Kepp
Birgit Schiøtt
Rebecca C. Wade

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF PROTEIN COMPLEXES IN THE BLOOD COAGULATION CASCADE

by

Jesper Jonasson Madsen

A dissertation presented to the PhD school DTU Chemistry of the
Technical University of Denmark in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

August 2014
Kgs. Lyngby, Denmark

Preface

This dissertation concludes my industrial PhD-project carried out at the Department of Chemistry, Physical and Biophysical Chemistry, Technical University of Denmark and Haemophilia Biochemistry, Novo Nordisk A/S under the supervision of associate professor Günther H. Peters, *PhD* (DTU) and principal scientist Ole H. Olsen, *PhD* (Novo Nordisk A/S) as an external advisor. The bulk of the presented dissertation is centered at studying the two central factor Xase (tenase) complexes of the blood coagulation cascade using computational biophysics and, in particular, the molecular dynamics (MD) method.

At the very core of the performed studies lies an attempt to understand function on the basis of structure. *I.e.* if we know the structure of a complex, can we elucidate its functional characteristics? The blunt and conservative answer to this question that has arisen over time is: *sometimes*. The reason for this is that many functional properties are in many cases a manifestation of dynamical behaviour and, hence, only poorly represented static structures. Enter MD. While there are many technical caveats and inherent limitations of present implementations of MD, the *in silico* method has proven itself a valuable tool over the last > ~30 years, in particular when used in conjugation with other complementary techniques (often wet-lab experimental ones).

The research have been performed primarily at Novo Nordisk A/S and using the affiliated supercomputer named “Davinci” to execute the MD simulations. Though, during my stay in the spring of 2013 as a visiting scientist with the Tajkhorshid group at the Beckman Institute, University of Illinois at Urbana-Champaign, I also exploited local resources for our joint study of the membrane-interaction of factor VIIIa.

The dissertation is structured as follows. Starting with a general introduction of the biochemical components (Chapter 1) and the applied methodologies (Chapter 2), the major parts of this thesis consist of manuscripts of scientific papers (Chapters 3 & 4 contain manuscripts (ms.) 1-5, of which ms.2, ms.3, and ms.4 have been submitted and are under

peer review) that are rounded-off and complete entities in themselves. Concluding remarks are provided in Chapter 5.

Kgs. Lyngby, August 2014

Jesper Jonasson Madsen, *M.Sc.Eng.*

Table of Contents

Summary

In this dissertation, components of coagulation factor (F)X-activating (tenase; fXase) complexes are studied using computational biophysics and the molecular dynamics (MD) method. The main focus is on structure-function relationships of the two central tenase complexes with respect to activation processes, complex formation, and platelet membrane association.

Both the intrinsic and the extrinsic tenase complex consists of a trypsin-like serine protease and auxiliary domains complexed with the appropriate cofactor; FIXa with FVIIIa and FVIIa with tissue factor (TF), respectively.

Topics covered will include the structural and dynamical changes upon proteolytic activation and TF-induced allosteric activation of FVIIa. In addition, FVIIa variants with the 170-loop grafted from trypsin will be looked into and, in particular, the mechanisms that enable these variants to have increased activity without TF are explained. Furthermore, the inter-domain linker connecting the two epidermal growth factor-like domains of FVIIa will be discussed with respect to consequences for its ability to form a productive complex with tissue factor. Finally, membrane binding of FVIIIa as mediated by the tandem C2-like domains is described using a highly mobile membrane-mimetic model.

Dansk resumé

Denne afhandling omhandler anvendelse af beregningsmæssige biofysiske metoder, specielt molekyldynamik, til karakterisering af koagulationsfaktor (F)X-aktiverende (FXase; tenase) protein komplekser. Primær fokus vil være på struktur-funktion relationen, der gælder for de to centrale tenase-komplekser i forhold til aktiveringsprocesser, kompleksdannelse, og interaktionen med blodpladememembraner.

Både det intrinsiske og det extrinsiske tenase kompleks består af en trypsin-lignende serinprotease med tilhørende domæner, som kræver en passende cofaktor; FIXa med FVIIIa og FVIIa med vævsfaktor (TF), respektivt.

Emner, der vil blive berørt i denne afhandling, inkluderer de strukturelle og dynamiske ændringer efter proteolytisk aktivering of TF-induceret allosterisk aktivering af FVIIa. Derudover kigges der nærmere på FVIIa-varianter hvor 170-løkken fra trypsin er integreret og, i særdeleshed, de mekanismer, som giver disse varianter øget aktivitet uden TF. Desuden vil interdomæne-linkeren, som forbinder de to epidermal vækstfaktor (EGF)-lignende domæner i FVIIa blive diskuteret i forbindelse med konsekvenser for dens medbestemmende rolle i at tillade en produktiv interaktion med TF. Slutteligt kigger vi nærmere på membran-interaktionen af FVIIIa, som er medieret af de to tandem C2-lignende domæner, ved brug af en mobil membran-mimetisk model.

Acronyms, abbreviations and terms

AMBER, *Assisted Model Building with Energy Refinement*

APC, *Activated protein C*

ASIS, *Active-site inhibited (coagulation factor) seven (by FFR-CMK)*

AT, *Anti-thrombin*

Å, *1 Ångstrøm = 0.1 nanometer*

BHK, *Baby hamster kidney*

C-terminus, *Carboxyl-terminus*

CHARMM, *Chemistry at Harvard Molecular Mechanics*

CHO, *Chinese hamster ovary*

CMAP, *φ-, ψ- cross term map correction for CHARMM*

CR, *Complement-type repeats*

CTR, *Chymotrypsin*

Da, *1 Da = 1 u = 1.660538921(73)×10^{-27} kg*

Discoidin, *Factor V/VIII C2-like*

DTU, *Technical University of Denmark*

EGF, *Epidermal growth factor*

FEM, *Feature-enhanced electron density map*

FF, *Force field*

FFR-CMK, *D-Phe-Phe-Arg-chloromethyl ketone*

FVII, *Coagulation factor VII*

FVIIa, *Coagulation factor VII (activated)*

FVIIIa, *Coagulation factor VIII (activated)*

FVIIIdB, *B domain-deleted coagulation factor VIII*

FIXa, *Coagulation factor IX (activated)*

FX, *Coagulation factor X*

FXa, *Coagulation factor X (activated)*

Gla, *γ-carboxyglutamic acid-rich*

GROMOS, *Gröningen Molecular Simulation*

HDX-MS, *Hydrogen-deuterium exchange mass spectrometry*

HMMM, *the Highly Mobile Membrane-mimetic Model*

HOMO, *Highest occupied molecular orbital*

HSPG, *Heparin sulphate proteoglycans*

k_{cat}, *Catalytic constant*

K_m, *Michaelis constant*

KNCO, *Potassium cyanate*

LDL, *Low-density lipoprotein*

LUMO, *Lowest unoccupied molecular orbital*

LRP, *Lipoprotein receptor-related protein*

MASP-1, *Mannan-binding lectin serine protease-1*

MD, *Molecular dynamics*

MEM, *Maximum-entropy electron density map*

MM, *Molecular mechanics*

N-terminus, *Amino-terminus*

NN, *Novo Nordisk A/S*

OPLS, *Optimized Potentials for Liquid Simulations*

PAR, *Protease-activated receptor*

pABA, *Para-aminobenzoic acid*

PC, *Protein C*

PDB, *Protein Data Bank*

PDF, *Probability density function*

PME, *Particle mesh Ewald*

PROPKA, *Heuristic pKa estimation*

PT, *Prothrombin*

RAP, *Receptor-associated protein*

RMS, *Root mean square*

RMSD, *Root mean square deviation*

RMSF, *Root mean square fluctuation*

TF, *Tissue factor*

TFPI, *Tissue factor pathway inhibitor*

S1-S4, *Spike 1 – Spike 4*

S2288, *H-D-Ile-Pro-Arg-pNA·2HCl (chromogenic substrate)*

Serpin, *Serine protease inhibitor*

SP, *Serine protease*

SPR, *Surface plasmon resonance*

sTF, *Soluble tissue factor; residues 1-219*

SVM, *Support vector machine*

tenase, *factor Xase*

tenase complex (intrinsic), *FVIIIa:FIXa*

tenase complex (extrinsic), *TF:FVIIa*

UB, *Urey-Bradley*

UIUC, *University of Illinois at Urbana-Champaign*

vdW, *van der Waals*

vWF, *von Willebrand Factor*

Xtal, *X-ray crystallographic*

List of figures

List of tables

1. General introduction

1.1 Haemostasis and haemophilia

In this chapter, we shall introduce the biomolecular components and mechanisms of relevance to this dissertation into a broader biological context.

In the blood circulates the molecular apparatus to repair traumatic injury and stop bleedings (up until a certain point). Haemostasis is the name for the complex network of inter-connected biochemical reactions which occur upon vessel injury and, when functioning properly, stop the bleeding. The haemostatic system has many components and an exhaustive review of the involved molecular biology is out of scope for the present text. An impressively thorough overview of principles in haemostasis and thrombosis can be found in the book by Marder *et al.*(Marder, Aird, Bennett, Schulman, & White, 2012)

For our purpose, however, a relatively brief and idea-driven introductory description of haemostasis is going to suffice. The two principal processes of haemostasis are:

1. The adherence of thrombocytes (platelets) to sub-endothelial collagen.
2. Stabilization of the thrombocyte matrix by a dense cross-linked network of fibrin.

The first step involves the glycoprotein-complex receptor GPIb-V-IX, which binds von Willebrand Factor (vWF), thereby allowing the platelet adhesion and plug formation at the spot where the blood vessel has been injured. The following step stabilizes the thrombocyte plug by a highly dense and cross-linked fibrin network. The biochemical reactions governing this second step is collectively designated the blood coagulation cascade and they involve many biomolecular components (coagulation factors), who interact promptly in a complicated yet highly regulated manner (Fig. 1). (MacFarlane, 1964)(Davie & Ratnoff, 1964) Two pathways, the extravascular (*extrinsic*) and the intravascular (*intrinsic*) both lead to the activation of the zymogen factor X (FX), which, in turn, combines with its enzyme cofactor, factor Va, to form the prothrombinase complex responsible for the rapid generation of thrombin. Thrombin is a key player in haemostasis

because it is the only one of the haemostatic proteases that is able to covert fibrinogen into fibrin, which polymerizes in the presence of FXIIIa to make the clot.

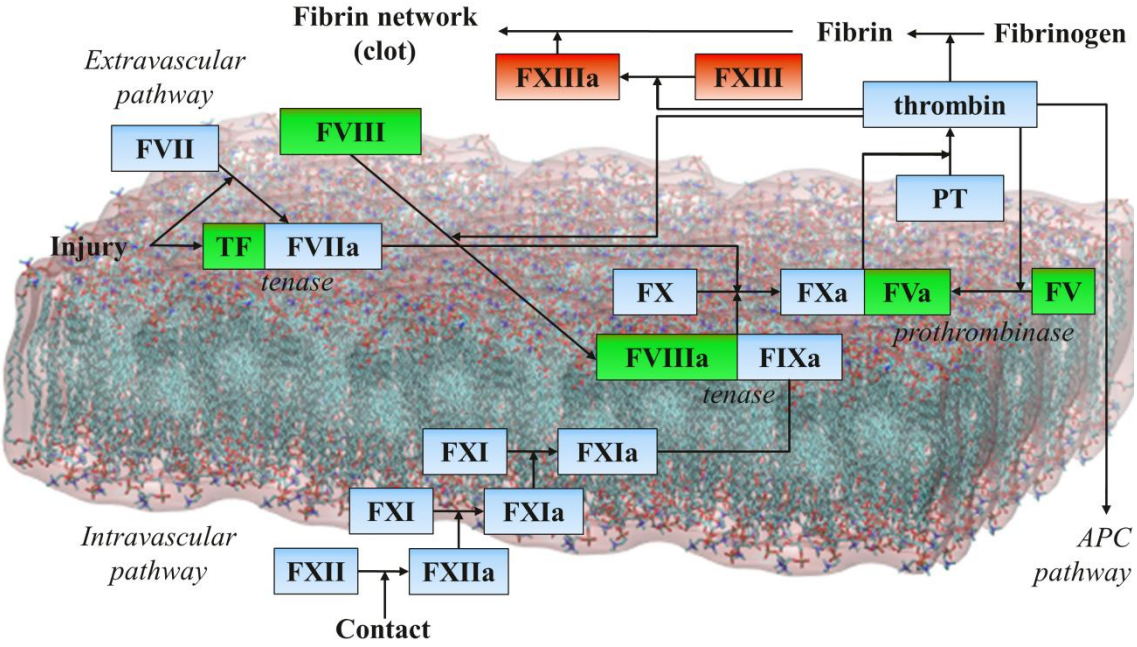


Figure 1 ***Schematic depiction of the coagulation cascade reactions leading to fibrin clot formation.***

The membrane is shown with sticks and half-transparent surface (pink). Haemostatic serine proteases ((pro)thrombin, FVII(a), FIX(a), FX(a), FXI(a), FXII(a)) are indicated by blue boxes and protease protein cofactors (TF, FV(a), FVIII(a)) by green boxes. FVIII(a), a transglutaminase, is shown in red. Explicitly labelled are the two tenase complexes (TF:FVIIa, the extrinsic tenase complex; FVIIIa:FIXa, the intrinsic tenase complex) and the prothrombinase complex (FXa:FVa).(Reconstructed from Fig. 7.1 of (Engelke, 2010))

Tissue Factor(TF)-triggered coagulation is initiated by the extrinsic tenase complex TF:FVIIa in the presence of Ca^{2+} and phospholipids. The intrinsic pathway involves the activated FVIII and FIX, which make up the membrane-bound intrinsic tenase complex that is potently capable of activating FX. Complex formation of the tenase complexes on the membrane surface provides a mechanism to localize the clotting of blood to the injury site.

Fig. 1 is deliberately kept simple and while reality *in vivo* is considerably more complicated, the most important reactions of the coagulation cascade are well represented

by the schematic. What is not elaborately shown in Fig. 1, however, is detailed co-factor dependence of each individual chemical reaction and in which ways coagulation is successively stopped. The latter is achieved mainly through the activated protein C (APC) anti-coagulation pathway, where the activation of PC depends on thrombomodulin and protein S, but also by the serine protease inhibitors (Serpins) tissue factor pathway inhibitor (TFPI) and anti-thrombin (AT). APC inactivates FVa and FVIIIa by proteolytic cleavage whereas TFPI inhibits FVIIa and FXa, and AT inhibits FXa and thrombin. A recent review of the protein C biology is found in (Mosnier, Zlokovic, & Griffin, 2007).

Individuals suffering from haemophilia A (B) are deficient in FVIII (FIX) and are, as a consequence, not able to activate enough FX to get proper amplification of the coagulation signal, negatively affecting the bloods ability to clot when needed. Roughly 400,000 males around the world suffer from haemophilia of which only ~25 % are currently receiving treatment according to ("World Federation of Hemophilia," 2014). Disease symptoms vary and include prolonged bleedings, joint bleedings, muscle and sub-cutaneous bleedings, internal bleedings, and intra-cranial bleedings. The typical treatment for haemophiliacs involves replacement therapy where the missing coagulation factor is provided. E.g. dosing with a recombinant (r)FVIIa, such as NovoSeven®, can provide haemostasis in patients with severe haemophilia A or B, offering a safe treatment without significant side-effects in most patient cases.

The coupling between structure and function is necessary for basic understanding at the molecular level of theory. Furthermore, such insights can lead to the development of therapeutic pro- and anti-coagulant treatments to benefit for patients with haemophilic and thrombotic disease, respectively.

1.2 Molecular structures and mechanisms in coagulation

1.2.1 Serine proteases and cofactors

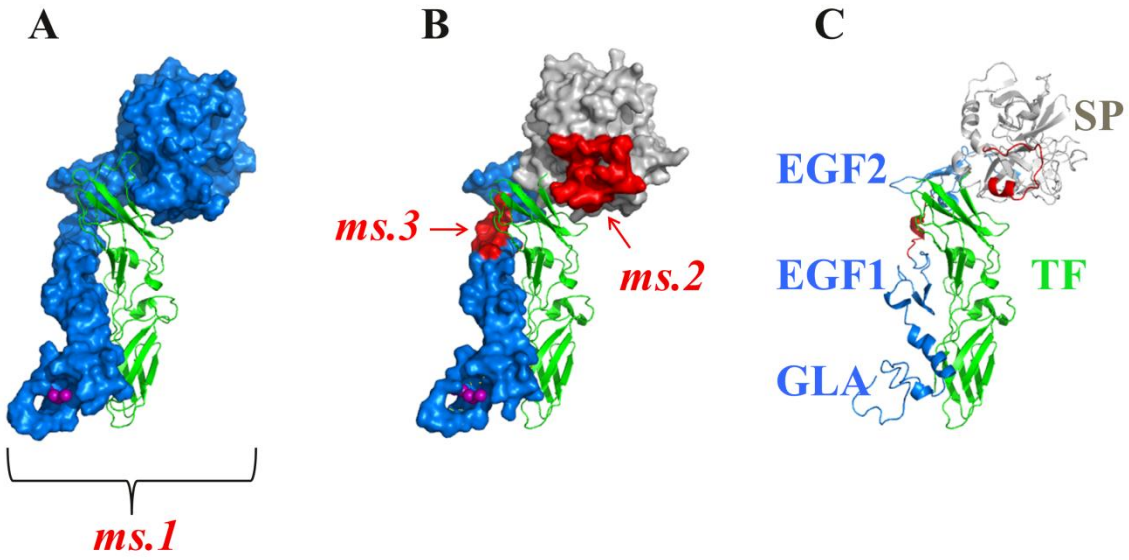


Figure 2 *Graphical abstract for manuscripts 1-3 regarding the* extrinsic *tenase complex.*

*(**A**) Crystal structure of the FVII zymogen structure (ms.1). FVII is shown as surface (blue) and TF in wire representation (green). The bound Ca^{2+} ions are shown as spheres (purple). (**B**) The structure of the TF:FVIIa tenase complex (PDB entry code 1dan*(Banner et al., 1996)*) with the TF-binding helix (ms.2) and the EGF1-EGF2 inter-domain linker highlighted (ms.3). FVIIa light chain (blue), FVIIa heavy chain (gray) is shown as surface and TF in wire representation (green). The bound Ca^{2+} ions are shown as spheres (purple). (**C**) The structure of the TF:FVIIa tenase complex (PDB entry code 1dan*(Banner et al., 1996)*) shown in wire representation including domain nomenclature for FVIIa.*

The precursor FVII (Fig. 2A) is produced as a single chain polypeptide molecule(Broze & Majerus, 1980). The architecture of FVII is a C-terminally positioned trypsin-like serine protease with three auxiliary domains; a vitamin K-dependent membrane-anchoring GLA-domain, and two epidermal growth factor (EGF)-like domains (Fig. 3). This exact domain template is shared among factors VII, IX, X, and protein C.

Figure 3 *Domain architecture of coagulation factors VII, IX, X, and protein C*

The concept of the C-terminally positioned serine protease is wide-spread with members represented begin responsible for maintaining or assisting many physiologically important processes including digestion, immune response, blood coagulation, fibrinolysis, and reproduction.(Hedstrom, 2002) For instance, other coagulation and fibrinolytic proteins with this generalized architecture include FXI, FXII, prothrombin, protein S, plasminogen, urokinase plasminogen activator, tissue-type plasminogen activator, and the success of such a design suggests a lego-block evolutionary process.(Patthy, 1985)

There is a preferred standard orientation of view for proteases to which we will adhere when visualizing molecular structures.(Gomis-Rüth, Botelho, & Bode, 2012) For FVIIa (Fig. 4A), this will correspond to aligning the active site cleft horizontally and having the non-prime side to the left and prime side to the right. This orientation positions the TF-binding helix and following 170-loop at the periphery of the SP domain to its left side with a top view of said helix (Fig. 4B).

Catalytic competency can be sampled (experimentally) at different levels of complexity (Table 1).(Mugherli, Burchak, Chatelain, & Balakirev, 2006)(Butenas, DiLorenzo, & Mann, 1997)(Morrissey, 2005) The trend is that for the larger substrates it gets increasingly difficult to infer the structural basis of the resulting measured signal. Small molecule and peptidic substrates probe the catalytic efficiency of the enzyme locally around the active site residues, the oxyanion hole, and the integrity of the S1 pocket, whereas the degradation of protein substrates potentially has major contributions from exo-sites and protein-protein interactions. Clotting assays are the coarsest grained and can be complex to interpret or rationalize. In ms.3 and subsequent notes on the 170-loop-swapped FVIIa variants, we shall use the amidolytic activity of FVIIa variants to gauge catalytic competency with and without TF.

Table 1 *Protease catalysis assays common for FVIIa and corresponding structural inference*

Substrate	Signal	Structural inference
Small molecule substrates, e.g. esters	Colorimetric or fluorometric	Oxyanion hole
Peptidic substrates (amidolytic activity)	Colorimetric or fluorometric	Oxyanion hole, S1 pocket, sub-sites specificity
Protein substrate (proteolytic activity)	FX activation (indirect)	Oxyanion hole, S1 pocket, sub-sites specificity, exo-sites specificity, protein-protein interactions
Thrombin assay	Thrombin-generation (indirect)	Complex
Clotting assay	Viscosity change	Complex

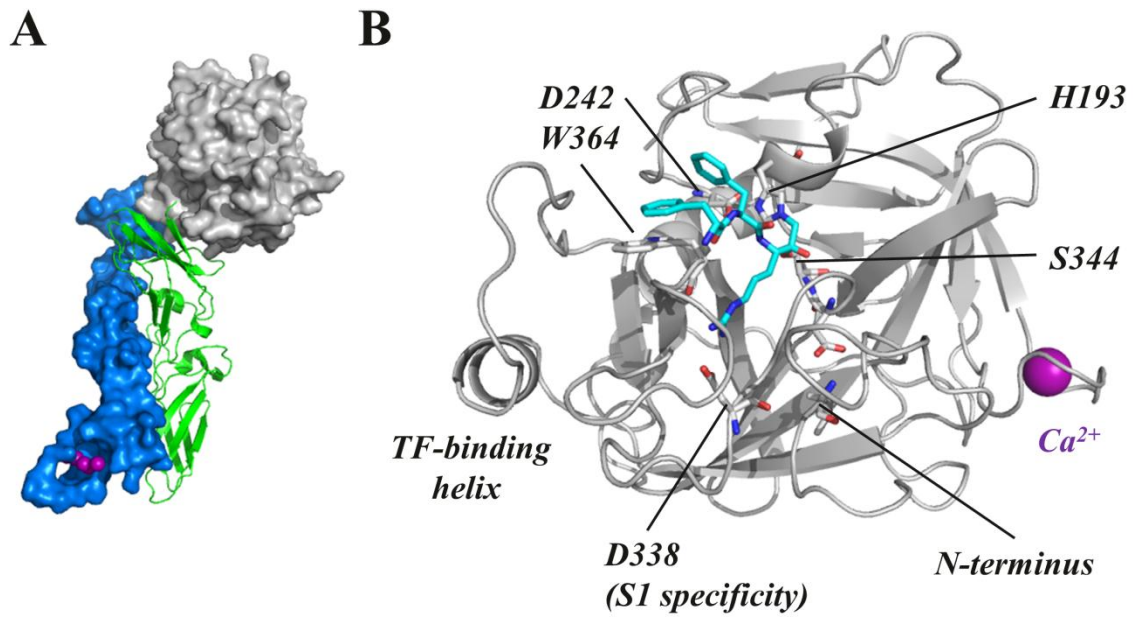


Figure 4 *The TF:FVIIa tenase complex and protease domain.*

*(**A**) Structure of the TF:FVIIa tenase complex based on PDB entry code 1dan*(Banner et al., 1996). *TF is shown in wire representation (green), the FVIIa light chain domains GLA, EGF1, EGF2 as surface (blue), and the FVIIa heavy chain SP domain as surface (gray). Ca^{2+} ions are shown as spheres (purple). (**B**) The FVIIa SP domain in wire representation and standard orientation with certain residues[1] explicitly shown as sticks (I153, H193, D242, S344, W364). The co-crystallized covalent inhibitor is shown as sticks (cyan). The bound Ca^{2+} ion is shown as a sphere (purple).*

FVIIa structures.

A comprehensive list of FVII/FVIIa/TF solved structures is maintained by James H. Morrissey.(http://www.tf7.org/) While these are generally well-represented by the by the structure of FFR-CMK-inhibited FVIIa (ASIS) with soluble tissue factor (PDB entry code 1dan, (Banner et al., 1996)) or the truncated EGF2-SP construct of FVIIa inhibited by Benzamidine (PDB entry code 1kli(Sichler et al., 2002)), certain structures stand out and warrant additional attention (Table 2). In particular, two TF-free structures of FVIIa with a visible EGF1 domain exist (PDB entry codes 1qfk(Pike, Brzozowski, Roberts, Olsen, & Persson, 1999)and 1dva(Dennis et al., 2000)) (Fig. 5).

---

[1] Two prevalent numbering schemes for FVII/FVIIa residues exist: Full-length FVII numbering and chymotrypsin-alignment numbering. Throughout this dissertation we the former are as standard and without explicit resignation, whereas chymotrypsin-numbering, when used, will be followed by ‘c’ in curly brackets, e.g. H193{c57} is the active site histidine-193 {ctr. res. no. 57}.

Table 2 *Survey of crystal structures of FVIIa with features distinct from the representative PDB entry code 1dan.*

PDB	Feature	Comment	Reference
1qfk	Free FVIIa with visible EGF1 domain.	Stabilized by crystal contacts. Gla domain not visible.	(Pike et al., 1999)
1dva	Free FVIIa with visible EGF1 domain.	Dimerically 'curled up'. Gla domain not visible.	(Dennis et al., 2000)
1klj	Uninhibited and free FVIIa.	Truncated construct containing EGF2 and SP. The co-crystallized inhibitor (benzamidine) was soaked out.	(Sichler et al., 2002)
2f9b	Alternative W364{c215} position/rotamer.	Bulky heterocyclic inhibitor with which W364 stacks.	(Rai et al., 2006)
2a2q	Alternative conformation of peptide bond Lys192-Gly193		(Bajaj, Schmidt, Agah, Bajaj, & Padmanabhan, 2006)

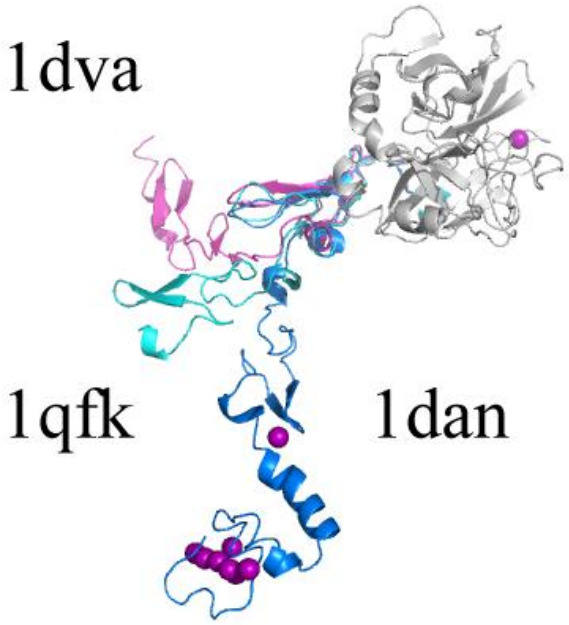


Figure 5 *Comparison of available structures of free FVIIa with a visible EGF1 domain to the extended conformation of FVIIa adopted when complexed with TF.*

Shown in wire-representation are PDB entry codes 1dan('H', heavy chain in gray; 'L', light chain in marine blue), 1qfk('L', light chain in cyan), 1dva('L', light chain in pink). The three structures have been structurally aligned based on their respective SP domains ('H', heavy chains).

It is evident that the light chain of free FVIIa is able to co-exist in (at least) several a couple of major conformations (Fig. 5). Furthermore, because there is no clashes between the light and heavy chain, it seems plausible that the entire span of intermediate conformations are equally like to also co-exist in the dynamical ensemble, brining into question whether the predominant species of free FVIIa in solution is the extended form know from the TF:FVIIa complex. The free FVIIa structures are stabilized by the protein-protein contacts (1qfk contacts a crystal neighbour and 1dva dimerically 'curled up'), but their mere existence reveals that the observed light chain conformation are not un-physical.

FVIIa and sodium.

There is some evidence that ions besides Ca^{2+} could bind to FVIIa and therefore be considered cofactors under certain conditions. In particular, has been proposed that Mg^{2+} could take the place of three of the seven bound Ca^{2+} ions and, furthermore, that FVIIa contains two Zn^{2+} binding sites and a Na^{+} binding site(Bajaj et al., 2006). Na^{+}-induced allostery is of fundamental importance in for instance thrombin(Gandhi, 2010; Huntington, 2008) where its stimulating effects and underlying molecular mechanisms of allostery have been elucidated(Gandhi, Chen, Mathews, & Di Cera, 2008). However, the crystal structures of Bajaj and co-workers were obtained under salt concentrations far above the physiological relevant range which, combined with the lack of stimulating effect of Na^{+} in FVIIa in the presence of Ca^{2+}, brings into question the potential significance of a putative role for Na^{+} in FVIIa allostery.

The serine protease hydrolysis mechanism.

A recent and rather comprehensive review of the serine protease mechanism and specificity can be found in (Hedstrom, 2002).

The intermediary structures formed in between enzyme and substrate in the hydrolysis mechanism of a serine protease has four principal states: The resting enzyme (Fig. 6), (tetrahedral) intermediary I (Fig. 7), the acyl-enzyme complex (Fig. 8), and the (tetrahedral) intermediary II (Fig. 9). The enzymatic mechanism proceeds as follows. Initially, a nucleophilic attack by the active site serine (195{c}) is performed in the peptide carbonyl. The nucleophilic property of said serine is enhanced due to stabilizing interactions with the active histidine, which, in turn, is held in a “protein acceptor state” by an aspartate (D102{c}).

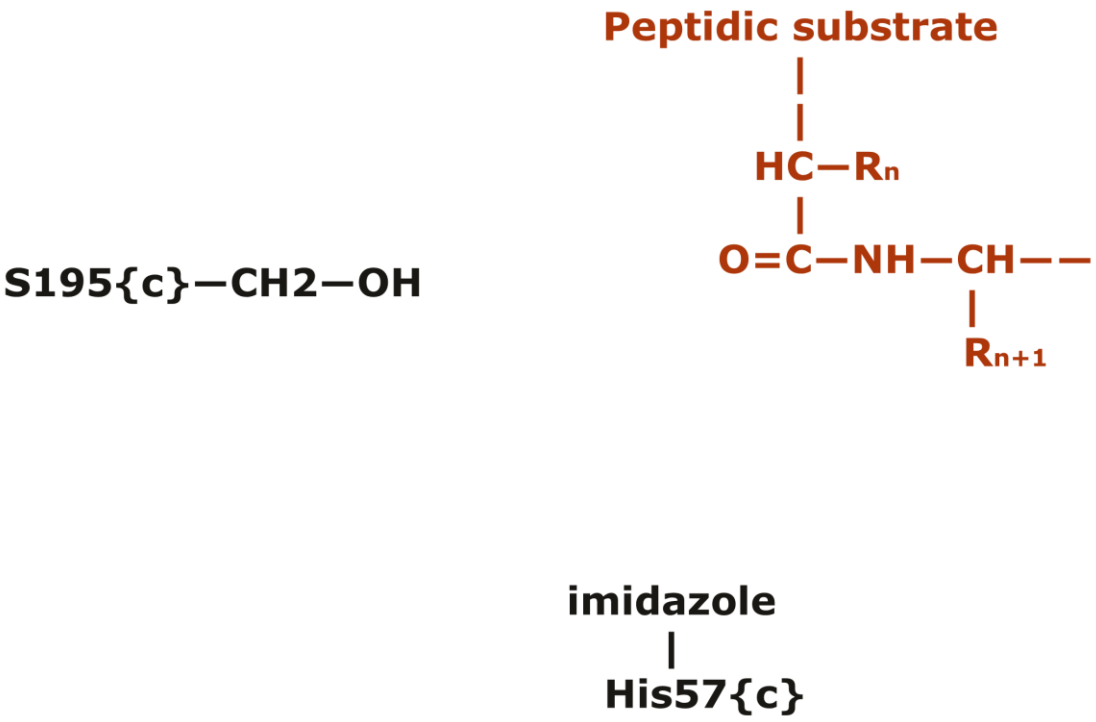


Figure 6 *The resting state of the catalytic serine and histidine residues of the serine protease.*

The oxygen anion in the tetrahedral intermediary I (and II) is stabilized by hydrogen bonding to the NH groups of S195{c} and G193{c} in the enzyme's so-called oxyanion hole (Fig. 10). The next step of catalysis is the reconstruction of the carbonyl double bond while releasing the C-terminal peptide leaving group. This is the step at which the peptide substrate is cleaved and it results in the acyl-enzyme complex where the N-terminal peptide is bound as an ester to S195{c}.

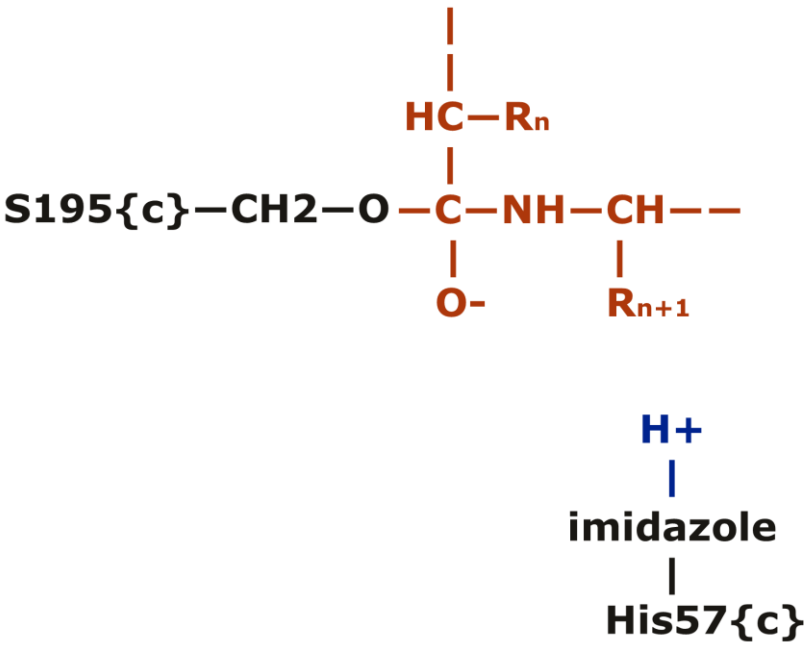


Figure 7 *The first tetrahedral intermediary structure of the enzymatic mechanism of the serine protease.*

The acyl-enzyme complex is hydrolysed in the deacylation reaction and this conversion consumes a water molecule to produce the tetrahedral intermediary II, where, as before, the oxygen anion is stabilized in the oxyanion hole until the N-terminal peptide is released and the enzyme active site is restored to its resting state.

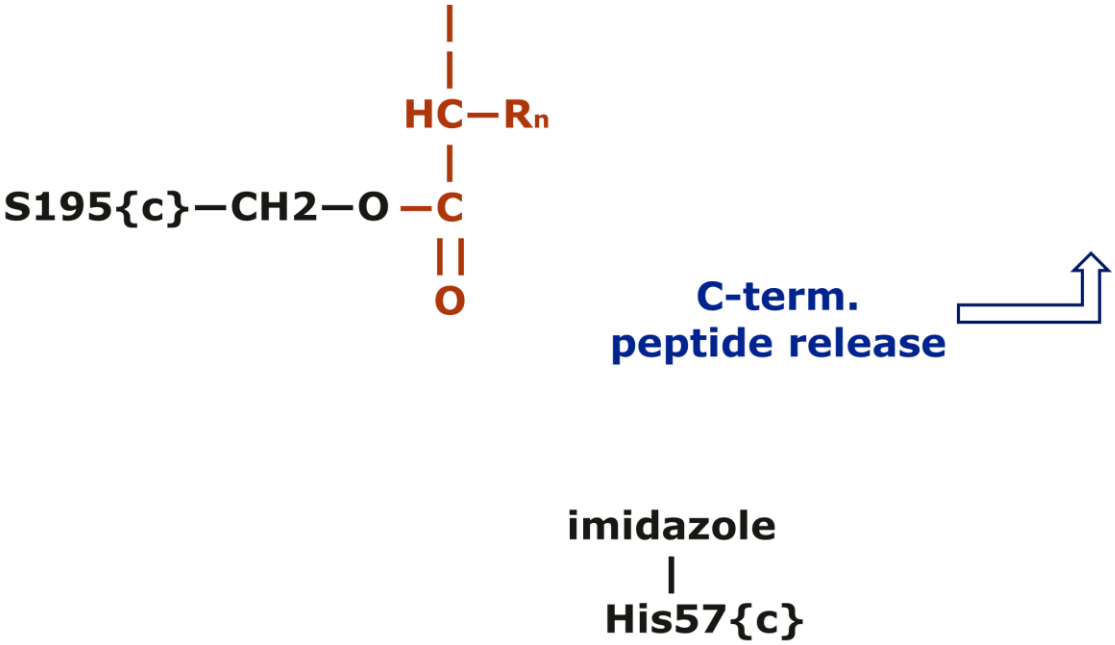


Figure 8 *The acyl-enzyme complex structure of the enzymatic mechanism of the serine protease.*

The organic chemist might wonder if the mentioned tetrahedral intermediates (Figs. 7 & 9) indeed are formed or the reaction actually follows the S_N2 reaction pattern in a single step, which you might expect from alkyl substrates (halides e.g.). Indeed, this topic has been controversial in the early years of investigation, however, we now know from molecular orbital theory that acyl and alkyl substrates have very different electronic orbital constructions. In particular, differences in the lowest unoccupied molecular orbitals (LUMOs) of the electrophilic substrates, which accepts nucleophilic attack of the highest occupied molecular orbital (HOMO) from S195{c}, satisfactorily explains why the scissile peptide bond cannot be broken directly in the enzymatic mechanism. Further structural subtleties in the serine protease mechanism has been elucidated recently including a ring-flip of H57{c}.(Ash et al., 2000)

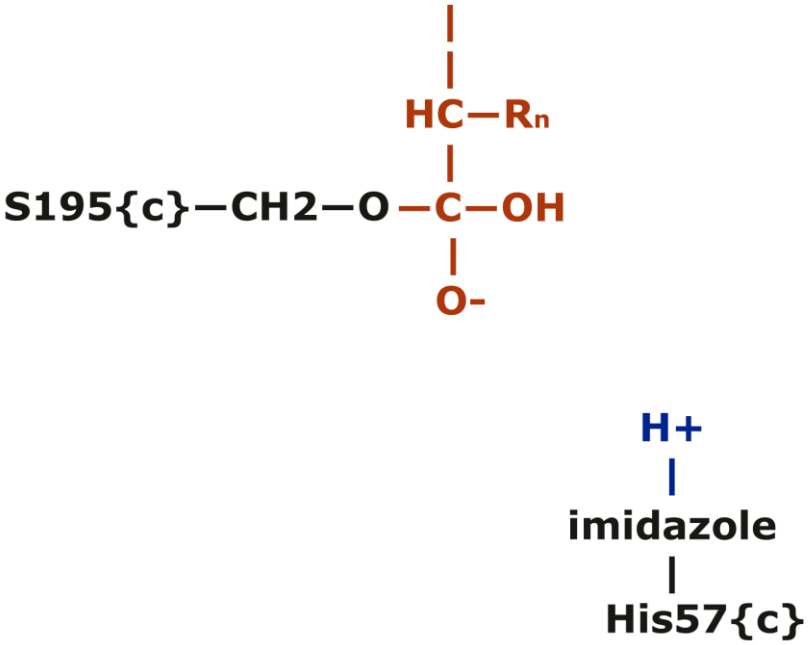


Figure 9 *The second tetrahedral intermediary structure of the enzymatic mechanism of the serine protease.*

A primary determinant of substrates specificity of serine proteases is the sidechain of the residue R_n (using the nomenclature from the above figures; also called position P1) because this moiety interacts directly with the enzyme by inserting into a substrate pocket, the sub-site S1 (Fig. 10).(Hedstrom, 2002) In FVIIa, that is known to have trypsin-like specificity, D189{c} resides in the bottom of the S1 pocket, consistent with its preference for R or L at position P1 of the peptide substrate (Fig. 10). By comparison, chymotrypsin has a more hydrophobic S1 pocket (incl. S189) and prefers F, Y, or W at position P1.

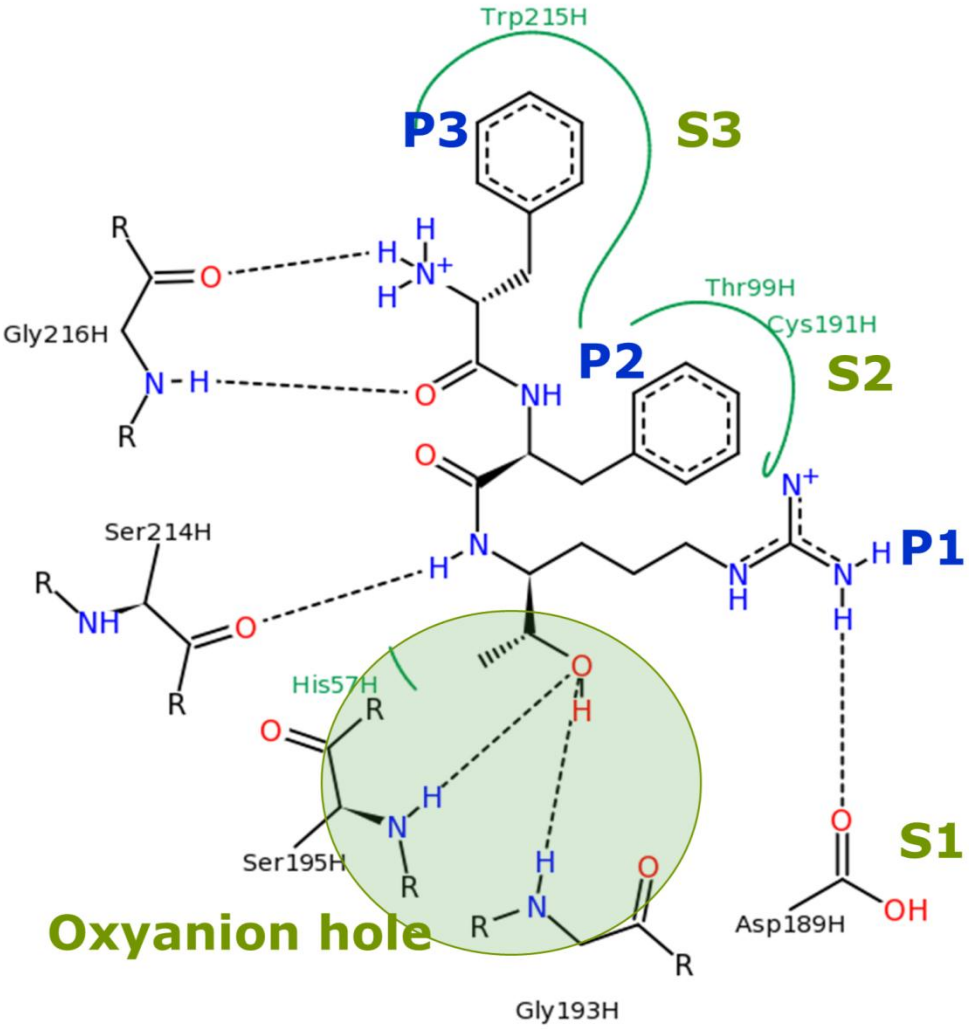


Figure 10 *Interaction between FFR-CMK inhibitor and FVIIa in ASIS indicating the position of enzyme sub-sites S1,S2, and S3, residue positions P1, P2, and P3 of the substrate (non-prime side), as well as the oxyanion hole.*

Interaction-figure of inhibitor and enzyme of ASIS (PDB entry code 1dan(Banner et al., 1996)) was generated using PoseView(Stierand, Maass, & Rarey, 2006). Residues are numbered according to chymotrypsin.

The prothrombinase complex.

A crucial step in the coagulation cascade reactions is the burst of thrombin produced by the prothrombinase complex as noted in the previous section 1.1. The FVa is cofactor to the serine protease FXa and on the (endothelial or platelet) membrane surface they assembly into the highly potent activator of prothrombin. Prothrombin activation proceeds by cleavage at two distinct sites (after R320 and R271) ensuing fibrin clot formation. The

prothrombinase complex, while important by itself, is of special interest to the work in this dissertation (Fig. 11A; ms.4) due to its homology with the intrinsic tenase complex (Fig. 11C), whose structure, to this date, has not yet been solved.

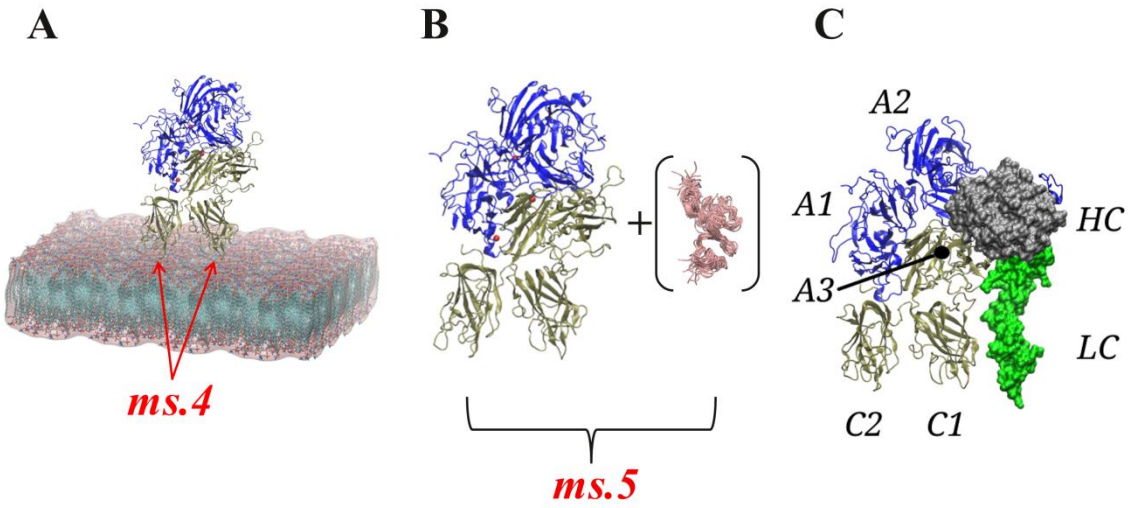


Figure 11 *Graphical abstract for manuscripts 4-5 regarding the* intrinsic *tenase complex.*

*(**A**) Membrane-interaction of FVIIIa discoidin domains (ms.4). (**B**) Clearance of FVIIIa by LRP1 (ms.5). (**C**) A putative FVIIa:FIXa tenase complex model.*

Recently, the re-engineered prothrombinase complex made up of proteins from the eastern brown snake (pseudonaja textilis) was solved.(Lechtenberg et al., 2013) This is a substantial advance toward understanding also the human counterpart and, by homology extension, possibly the FVIIIa:FIXa tenase complex. Since this structure will the key source in constructing a putative FVIIa:FIXa tenase complex model (Fig. 11C, ms.4), it warrants thorough examination.

In the snake venom prothrombinase complex, FXa snuggles in between the A2 and A3 domains of FVa. Surprisingly, a peptidic segment from FVa extends over the FXa SP domain and thereby embracing it. The positioning of FVa and FXa in the complex reveals that the FVa and FXa molecules roughly assume the equal relative (to the membrane) heights. The predominant role of the A2 the interaction between protease and cofactor

beautifully explains why removal of A2 from FVa by APC leads to dissociation of the prothrombinase complex and subsequent cessation of thrombin generation in the negative feedback loop (Fig. 1). While the interaction between prothrombin and the prothrombinase complex and cleavage of the substrate may not be directly evident, one could speculate on the position of prothrombin based on the established fact that prothombin is first cleaved at R320 (forming meizothrombin) and subsequently cleaved at R271 to release the mature thrombin molecule(Orcutt & Krishnaswamy, 2004), as indeed Lechtenberg, Murray-Rust and co-workers have done in their pioneering paper (Fig. 3 of (Lechtenberg et al., 2013)). Of further notice for the work presented in this dissertation is the observation that the snake venom FVa from the prothrombinase complex (PDB entry code 4bxs) has a significantly different relative orientation of FV C1 and FV C2 when compared with other (free) FVa structures (e.g. the APC-inactivated bovine FVa, PDB entry code 1sdd(Adams, Hockin, Mann, & Everse, 2004)), indicating mobility of the discoidin C2-like domains in the cofactor. In the former structure, the two consecutive tryptophan residues positioned in the FV C2 spike 1 are translated almost 30 Å in an out-of-plane back-kicking structural change relative to their respective positions in the latter structure (Fig. 12).

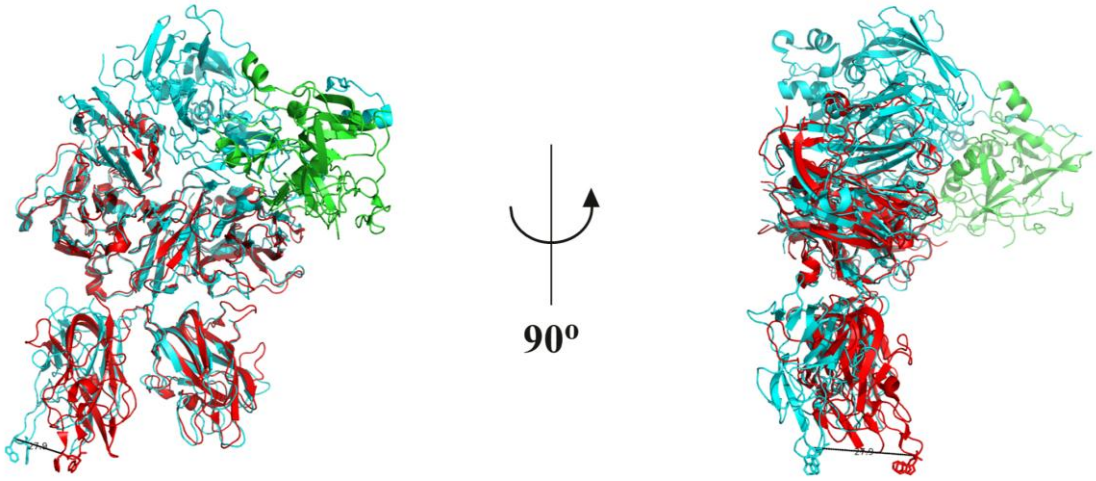


Figure 12 *Overlaid FVa molecules with different relative orientations of the tandem discoidin domains. Structural alignment of the snake venom prothrombinase complex in wire-representation (FVIa in cyan; FXa in green) from PDB entry code 4bxs*(Lechtenberg et al., 2013) *with the APC-inactivated bovine FVa in wire-representation (in red) from PDB entry code 1sdd*(Adams et al., 2004) *in standard orientation, same as Fig. 11C, (left panel) and a +90 degree rotated view around the z-axis (right panel). Highlited (as sticks) are the two adjacent tryptophans residing within the spike 1 of the C2 domains and a bar (length 27.9 Å) indicating their translated distance.*

FVIII catabolism.

The clearance of FVIII is known to be mediated by low-density lipoprotein receptor-related protein (LRP) (Fig. 11B) and glycoproteins, e.g. heparin sulphate proteoglycans (HSPGs). It is hypothesized that the initial interaction between the cell membrane and the FVIII/vWF complex happens primarily via HSPGs and then, once bound, FVIII can be taken up by LRP-mediated endocytosis.(Saenko, Shima, & Sarafanov, 1999)

In the next chapter, we will re-visit fundamental aspects of the molecular dynamics methodology.

2. Computational biophysics and the molecular dynamics method

The molecular dynamics (MD) method is a powerful technique for generating a dynamical ensemble, which with recent year's advancement in software and hardware capacities allows the study of fundamental biological systems and processes at relevant size and time scales. Mind you, not all time scales are equally accessible once the level of theory have been specified since, obviously, achievable simulation time and degree of detail relate inversely for a fixed amount of computational resources.

The time evolution (dynamics) of the modelled system is propagated by solving its equation of motion. To the degree that our systems can be well represented by classical mechanics, the following equation of motion applies

$$\vec{F}(\vec{r},t) = m\frac{d^2\vec{r}}{dt^2} = m\vec{a} = \nabla U$$

where $\vec{F}$ is the net force acting on the particle, m is the mass, and $\vec{a}$ is the acceleration vector describing the resulting kinematic motion. This above equation is of course nothing other than the familiar Newton's second law of motion.

The present chapter on MD methodology is not meant to be comprehensive. Instead, the interested reader is referred to classical texts such as (McCammon & Harvey, 1988) and (Allen & Tildesley, 1989), or modern incarnations hereof(Frenkel & Berend, 2001)(Rapaport, 2004).

2.1 The force field

Common empirical force fields for simulating biomolecular systems include CHARMM, AMBER, OPLS, and GROMOS. These FFs share the same basic structure in that contributions of the potential energy function, U, is divided into bonded and non-bonded

terms. The former are heuristic descriptions of the potential energy of bonded interactions and in CHARMM(MacKerell et al., 1998) they include terms to account for bond stretching, angles bending, torsion (or dihedral) rotation, improper (out-of-plane) bending, and a Urey-Bradley (UB) component (a cross-term for angle bending using 1,3 non-bonded interactions). The terms for bonds, angles and dihedrals are intuitive descriptors. The purpose of the UB term is to couple atoms which are two covalent bonds apart, and the improper term governs planarity of conjugated systems such as aromatic rings. Furthermore, a correction term for backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles (CMAP(Best et al., 2012)) is included.

The non-bonded interaction of the FFs are divided into two terms, U_{vdW} and $U_{electrostatic}$, representing long range dispersion interactions and electrostatic interactions, respectively. Of primary relevance to these terms are pairs of atoms not covalently bonded; non-bonded terms for these interaction pairs are typically scaled or discarded.

Mathematically, the potential energy function for the CHARMM (with CMAP) all-atom MD simulations can therefore be written as

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{improper} + U_{UB} + U_{CMAP} + U_{vdW} + U_{electrostatic}$$

where individual terms may be explicitly expressed as

$$U_{bond} = \sum_{bonds} K_{ij}^{b}(r_{ij} - r_{eq,ij})^2$$

$$U_{angle} = \sum_{angles} K_{ijk}^{a}(\theta_{ijk} - \theta_{eq,ijk})^2$$

$$U_{dihedral} = \sum_{dihedrals} K^{d}(1 + \cos(n\varphi - \delta))$$

$$U_{improper} = \sum_{impropers} K^{i}(\omega - \omega_{eq})^2$$

$$U_{UB} = \sum_{Urey-Bradley} K^{UB}(b^{1-3} - b_{eq}^{1-3})^2$$

$$U_{CMAP} = \sum_{residues} u_{CMAP}(\Phi, \Psi)$$

$$U_{vdW} = \sum_{i<j,vdW} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

$$U_{electrostatic} = \sum_{i<j,elec} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}}$$

The bond-, angle-, UB-, and improper terms are harmonic potentials with force constants K and equilibrium values for the respective (internal) coordinates denoted by subscript index eq. The distance between atoms i and j is denoted r_{ij}, while θ and φ denote angles defined by groups of atoms. The dihedral term is described by a cosine function with phase δ and number minima n. The van der Waals interaction is approximated by a standard 12-6 Lennard-Jones potential with well-depth ε_{ij} and collision distance σ_{ij}, which can be expressed as $\sigma_{ij} = \frac{\sigma_{ii}+\sigma_{jj}}{2}$ and $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$ (Lorentz-Berthelot combination rules). The electrostatic interaction is described by a Coulombic potential with partial charges on atoms i and j are q_i and q_j. ε_0 is the vacuum permittivity and ε_r is the relative dielectric constant (where the latter is set to unity in all-atom MD simulations because screening effects are directly accounted for). The set of parameters $\{K, r_{eq}, \theta_{eq}, n, \delta, \omega_{eq}, b_{eq}, \sigma, \varepsilon, q\}$ are fitted to against experimental thermodynamic properties and quantum mechanical calculations. For CHARMM, this fitting is performed in the explicit presence of water to optimize the description of aqueous systems (Mulliken procedure; (Mulliken, 1955)).

There are important limitations to this conceptually simple FF description of a physical system. Most notable, bonds cannot form or break, meaning that chemical reactions do not occur. Furthermore, the partial charges fixed and not reassigned during the simulations, so many-body effects such as (electronic) polarization is neglected.

2.2 The numerical integrator

The time step in the numerical integrator must be chosen to optimize efficiency while maintaining precision. On the formal side of the matter, the sampling theorem by Nyquist-Shannon(Shannon, 1998) tells us that if the highest frequency motion in the simulation is *W*, we should choose a time step of maximally *W/2* to adequately sample the motions.

Therefore, since our simulations will contain hydrogen atoms which have high-frequency bond vibrations(Nibbering et al., 2014), we are typically limited to pick time steps at around ~1.0 fs. However, we shall employ an algorithm to rigidify bond lengths, namely SHAKE(Ryckaert, Ciccotti, & Berendsen, 1977), by which we can roughly double the time step for the velocity Verlet integrator(Verlet, 1967).

2.3 Software

The MD software employed in the present work is called NAMD(Phillips et al., 2005) and based on charm++(Brooks et al., 2009). NAMD is a sophisticated tool and has several important features and functionalities aimed at computational accuracy and efficiency. However, describing these in details is out of scope for this text. A few include:

- Multiple Time Stepping, allowing long-range interactions to be computed less often
- Electrostatics calculations by the Particle-mesh Ewald algorithm(Essmann et al., 1995)
- Load balancing schemes and parallel scalability.

Visualization of molecular structures throughout this dissertation was made using either PyMOL(The PyMOL Molecular Graphics System, Schrödinger, LLC) or VMD(Humphrey, Dalke, & Schulten, 1996). Plots were prepared using Grace (xmgrace, http://plasma-gate.weizmann.ac.il/Grace).

2.4 Simulating the serine protease

The successful execution of MD simulations of biomolecular systems involves careful preparing of the systems. Of particular concern when simulating serine proteases is that the protonation state of the active sites residues (in chymotrypsin numbering H57{c}, D102{c}, and S195{c}) should adequately represent their most likely configuration, e.g. for physiological conditions the pH value should be ~7.

X-ray resolved crystallographic structures have the major limitation that hydrogen atoms cannot be seen. Furthermore, due to common ambiguity, the actual orientation of the imidazole ring of (the rotameric state) of the active site histidine is not known. Relevant protonation states for histidine include (in CHARMM language) HSE, which is neutral and the ε-nitrogen is protonated, HSD, which is also neutral but the δ-nitrogen is protonated

and, lastly, HSP, where both the ε-nitrogen and the δ-nitrogen is protonated. Besides the chosen protonation, also rotameric states could be considered.

Biochemistry textbooks usually cite the pKa of histidine (the imidazole sidechain) around ~6.0-6.5.(Stryer, Berg, & Tymoczko, 2002) However, estimates of the empirical tool PROPKA(Li, Robertson, & Jensen, 2005) for the FVIIa SP domain (PDB entry code 1kli(Sichler et al., 2002)) can be seen in Table 1. These results indicate that D102{c} is unprotonated (negatively charged) and H57{c} as doubly-protonated (positively charged). In addition, it the proton affinity of D189{c}, positioned in the bottom of the S1 specificity pocket, is predicted relatively high by PROPKAv.3.0.

Table 3 *PROPKA estimates of protonation states of selected residues of the FVIIa serine protease.*

Type	Res.No. {ctr.no.}	pKa (v.3.0)	pKa (v.3.1)	Model-pKa
His	344{57}	7.31	8.01	6.50
Asp	242{102}	4.39	3.58	3.80
Asp	338{189}	7.50	3.78	3.80

From left to right: Three-letter amino acid code for residue, residue number {chymotrypsin number}, PROPKA v.3.0 pKa estimate (protein-only), PROPKA v.3.1 pKa estimate (protein-ligand), reference value pKa. FVIIa structure was PDB entry code 1kli(Sichler et al., 2002).

Simulations at the level of theory applied in this dissertation do not allow for alternating protonation states, such as QM/MM hybrid methods(Born & Oppenheimer, 1927)(Car & Parrinello, 1985) or various heuristic constant pH schemes. Therefore, a single representative state must be picked. To represent a (uninhibited) state of the serine protease in which the H57{c} serves as the catalytic base in the deacylation step of hydrolysis (section 1.2 & Fig. 8), the H57{c} was neutral (CHARMM type HSE) and D102{c}+D189{c} were charged (-1 *e*, standard acidic titration state) though out. This is furthermore consistent with neutron diffraction experiments.(Kossiakoff & Spencer, 1981) Though, that the protonation states may become inadequate during the time course of the simulations, due to strong coupling between protonation and structural conformation, is of potential concern.

3. Extrinsic tenase complex

3.1 Proteolytic activation of FVII

3.1.1 Prelude

The draft manuscript 1 describes structure and dynamics of the FVII zymogen, a molecular structure that has remained elusive. The structure was based on a crystal of FVIII-S344A (active serine mutated to alanine) in complex with tissue factor which was discovered to be twinned. In addition, the solved structure contains relatively large unresolved segments for several loops, specifically around activation loops 1, 2, 3, and in the 170-loop.

To shed light in the molecular flexibility of these regions and the dynamics, in general, of the FVII zymogen, a homology model was constructed and studied.

Thermodynamics of complex formation. Equilibrium reactions of complex formation between FVII and FVIIa with TF are stoichiometric 1:1, i.e.

$$\underset{H_1,S_1}{\mathbf{FVII}} + \underset{H_2,S_2}{\mathbf{TF}} \xrightarrow{yields} \underset{H_3,S_3}{\mathbf{TF{:}\,FVII}}$$

$$\underset{H_4,S_4}{\mathbf{FVIIa}} + \underset{H_5,S_5}{\mathbf{TF}} \xrightarrow{yields} \underset{H_6,S_6}{\mathbf{TF{:}\,FVIIa}}$$

The enthalpic and entropic state functions $\{H_n, S_n\}$ are for each state of the molecules indicated. By definition, $H_2 = H_5$ and $S_2 = S_5$. If it is assumed that the state functions of FVIIa are in the same order as those of FVII, the major differences in TF-affinity between the activated and zymogen FVII originates from differences in terms $\{H_3, H_6, S_3, S_6\}$, i.e.

$$\Delta\Delta G \sim (H_6 - H_3) - T(S_6 - S_3).$$

This may seem like a crude approximation, but the zymogen-like property of FVIIa is well established on a functional level. The standard procedure used to quantify this (at least within the field of haemostasis-related biochemistry) is carbamylation(Hilgartner & Miller,

1975), where the N-terminus is chemically modified (by cyanate) to inactivate the enzyme. The assay is obviously indirect since the measured signal is residual enzymatic activity, but the conclusion holds unambiguously that TF protects the N-terminus of FVIIa from chemical modification and, by inference, promotes N-terminal insertion.

The major contributor to the first enthalpic term of the delta-delta Gibbs free energy (above equation) is the insertion of the N-terminus into the activation pocket, where it forms a stabilizing salt bridge with D242. The (signed) enthalpic gain of the formation of said isolated and buried salt bridge stems from favourable hydrogen-bonding and electrostatic interaction and amounts to approximately $-6.3(\pm 4.4)$ kcal/mol for typical salt bridges in proteins(Kumar & Nussinov, 1999). Considering that binding affinity of TF to FVIIa is some 5-fold enhanced over that to FVII, the lhs. amounts to

$$\Delta\Delta G \sim -RT \ln 5 = -(8.134 * 298.15 * \ln 5)\frac{J}{\text{mol}} = -0.9329\frac{\text{kcal}}{\text{mol}}$$

where R is the gas constant and T is the temperature. Thus, differences in entropic contributions (favouring the free form of the FVII zymogen) and enthalpic contributions (favouring the TF-complexed form of FVIIa) result in a subtle balancing of contributing terms. But they are likely large enough that experimental validation should be achievable, e.g. using isothermal titration calorimetry(Zakariassen & Sørlie, 2007).

3.1.2 Manuscript 1 (*preliminary*): Madsen JM, *et al*, in preparation

Draft

Crystal structure of the factor VII zymogen with soluble tissue factor elucidates the activation transition

Jesper J. Madsen[1,2], Lars Anders Svensson[3], Prafull S. Gandhi[1], Ole H. Olsen[1]

[1] Haemophilia Biochemistry, and [3] Protein Structure and Biophysics, Novo Nordisk A/S, Måløv, Denmark.

[2] Department of Chemistry, Technical University of Denmark, Kgs. Lyngby, Denmark.

Correspondence: Ole H. Olsen, Haemophilia Biochemistry, Novo Nordisk A/S, Novo Nordisk Park, F5.1.29, DK-2760 Måløv, Denmark.

Tel.: +45 3075 4511

E-mail: oho@novonordisk.com

Introduction

The solution of the X-ray crystallographic structure of factor (F)VIIa with soluble tissue factor (TF) marked a major breakthrough toward understanding the molecular mechanisms involved in the initiation of blood clotting(1). Ever since, however, a great desire has been to discover the structure of the FVII molecule in its (un-activated) zymogen form in order to better understand the activation transition undertaken. The activation of FVII involves proteolytic cleavage after position R152, positioned in between the serine protease (SP) domain and the epidermal growth factor-like 2 (EGF2) domain, splitting the molecule into two chains (heavy chain: SP; light chain: Gla, EGF1, EGF2) still covalently linked together by an inter-chain disulfide bridge, C135-C262.

Although the activation transition is, to some extent, analogous to that of trypsinogen-trypsin, which has been thoroughly studied, there are subtle yet important differences between FVII and trypsinogen. The first mechanistic description of the activation transition of trypsinogen emerged almost 40 years ago.(2) In trypsinogen, a hexameric peptide is removed from the N-terminal region upon activation to trypsin and this cleavage promotes the transition from a disordered state (trypsinogen) to an ordered state (trypsin).(3) It is well-known that intrinsically disordered peptidic segments of proteins lacking structure can carry the capacity to undergo disorder-to-order transitions upon e.g. ligand-binding. (4)(5) The defining functional distinction between FVIIa and trysin (as well as other homologous proteins) is that FVIIa, even once activated, is enzymatically incompetent and requires its biological cofactor, TF, for significant activity. This property is due to the an incomplete activation transition (or rigidification) of the activation domain of FVIIa(6)(7), which is popularly referred to as still being ‘zymogen-like’. Therefore, the actual nature (and structure) of the FVII zymogen is required to elucidate the activation process satisfactorily for this pivotal protease.

An eccentric activation mechanism of FVII involving reconfiguration of β-strands has previously been hypothesized(8). However, detailed experimental and theoretical investigations(9)(10)(11) do not support the reregistration hypothesis and it is not discussed further.

A further complication is the role of TF in the induction of order to the activation domain of FVII/FVIIa. The affinity maturation which occurs is bi-directional; substrate (or inhibitor) binding to FVII increases the affinity for TF and, vice versa, the TF:FVII complex has increased affinity toward substrates. A consequence hereof is that the rate of auto-activation of FVII is likewise increased by the binding of TF(12), but the mechanism through which this happens is not clear.

Materials and Methods

Protein production and X-ray crystallography. Crystals of the complex sTF:FVII-S344A was obtained by the hanging-drop method similar to previously described(13). Diffraction data were collected at beamline I911 MAX-lab, Lund, Sweden.(14)

Structural refinement.

The structure of TF:FVII-S344A was determined by molecular replacement based on the binary complex between FVIIa and TF (PDB entry code 1dan(1)). Automated refinement was performed using Refmac(15) and PHENIX(16) to a resolution of 2.7 Å with the twin law = h, – k, – l and twin fraction 0.5 (full twinning) for phenix_refine. Cycles of manual model building in Coot(17) into the feature-enhanced electron density map (FEM) continued as long as confident building into density (down to 0.5σ) with generally good geometry could be done. Refinement statistics can be found in Table 1.

Molecular Dynamics Simulations.

Model building.

There are several published zymogen structures of proteases with the chymotrypsin fold (Table 2). Retrieval of sequences and alignment of S1 peptidase regions/domains was done with Uniprot(18) and the clustal Omega algorithm(19). The missing loop segments in the FVII crystal structure were modeled using Quanta(Molecular Simulations Inc., San Diego, CA) based on trypsinogen resolved at 1.55 Å (PDB entry code 2ptn(20)). All bound Ca^{2+} ions and crystal water molecules were included in the constructed model.

Simulation details.

For the simulated systems (Table 3), all non-protein moieties present in the PDB files, except Ca^{2+} ions and crystal water molecules, were removed. The potential energy was initially minimized by applying the conjugated gradient method for 5,000 steps. At this point, the production run of length 100 ns was executed in the *NPT* ensemble with the target P = 1 atm and T = 310 K controlled by the Nosé-Hoover Langevin piston barostat(21)(22) and the Langevin thermostat (damping coefficient: 5/ps)(23), respectively. Throughout, anisotropic pressure coupling was applied (piston damping coefficient: 5/ps, piston period: 100 femtosecond (fs), and piston decay: 50 fs). Long-range electrostatic

forces were calculated using the Particle Mesh Ewald method(24)(25) with a grid spacing of approximately 1 Å and a fourth-order spline for interpolation. Electrostatic forces were updated every 4 fs. Van der Waals interactions were cut off at 12 Å in combination with a switching function beginning at 10 Å. Periodic boundary conditions were applied in *x*-, *y*-, and *z*-directions. The simulation was performed using the NAMD 2.7 software package(26) with the CHARMM27 force field(27), and the TIP3P water model(28). An integration time step of 2.0 fs was used for the velocity Verlet algorithm with SHAKE(29). All analysis and visualization of molecular structures was done using the Visual Molecular Dynamics (VMD) 1.9.1 software package.(30) Plots were prepared using Grace (xmgrace, http://plasma-gate.weizmann.ac.il/Grace).

Analysis.

RMSF. The Root mean square fluctuation (RMSF) of atomic positions for backbone atoms were calculated in a window width of 0.5 ns with a step size of 0.1 ns as

$$RMSF_{(v)} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2}$$

where T = no. frames

B factors.

X-ray crystallographic B factors relates to the rms fluctuation via the equation(31)

$$B_{(v)} \overset{\text{def}}{=} \frac{8\pi^2}{3} < MSF_{(v)} >$$

hence giving a quantitative measure for comparison between experiments and computation.

Model simulation fit into density. Cross-correlation coefficient at 5 Å resolution between the X-ray crystallographic electron density map and each frame of the FVII homology model simulation (SP domain only) were calculated using the MDFF plugin(32) in VMD(30).

Results and Discussion

The resulting determined FVII zymogen structure had missing loop segments at residues 144-156 (part of activation linker region), 286-294 (activation loop 1), 317-321 (part of 170-loop), 337-340 (part of activation loop 2), 364-373 (activation loop 3) (including the mentioned loop flanking residues). Autolysis sites are present in the 170-loop and activation loop 1 of FVIIa but no degradation is expected to have occurred because FVIIa-S344A is catalytically inactive.

No significant density (above noise level) was visible for the activation linker region and 170-loop residues in the feature-enhanced electron density map (FEM) or maximum-entropy electron density map (MEM). The flexible nature of the 170-loop is consistent with other structures of FVIIa, for instance the structure of a TF-free truncated construct of FVIIa containing only EGF2 from the light chain and where the co-crystallized inhibitor was soaked out(PDB entry code 1klj(33)). It is surprising, though, that the 170-loop is not stabilized (and hence visible in the crystal structure) in the presence of TF. There is no evidence of cleavage at R152, nor is I153 found in the activation pocket, confirming that the structure presented in this study indeed is the zymogen FVII. Residual density is observed in the FEM around the region where the activation loops 2 and 3 are positioned in FVIIa is observed. Intriguingly, these weak “tubes” of density (Fig. 5A) align not far from the main chain atoms in FVIIa. However, no confident model building based on this residual density could be performed with good geometry and the loop segments corresponding to activation loops 1,2, and 3 are possible predestined as too flexible in the absence of an inhibitor. This view is consistent with the disorder embraced by trypsinogen-like zymogens and likely further worsened by the extended loops in FVIIa.

The multiple sequence alignment guide tree generated by uniprot reveals that nearest sequence-based relatives among serine proteases with resolved zymogen structures are (in sorted order) FXI, MASP-1, and Thrombin. The structure of FXI(34), however, does not provide insight into the region around activation loops 1, 2, and 3, because these are, analogously to the present study’s FVII structure, also missing. Furthermore, structurally aligning MASP-1 SP domain(from PDB entry code 4igd(35)) with FVII results in a higher

overall geometric deviation (based on calculated RMS values) than the alignment of the structural ancestor trypsinogen with FVII. The model of the complete FVII zymogen was therefore constructed from the X-ray crystallographic structure with missing loops modeled based on trypsinogen(PDB entry code 2ptn(20)).

A major aim for the constructed homology model of the FVII zymogen is to approximate the structure in protein segments where no confident building could be performed guided by the X-ray diffraction data. FVII (and other coagulation serine proteases) circulate the blood as inactive zymogens and a hallmark structural modification is facilitated proteolytic activation and insertion of the newly formed N-terminus into the activation pocket, forming a salt bridge with D194. From a survey of published structures of zymogen proteases with the chymotrypsin family, that it is usual for the zymogen to have the same fold as the enzyme with the lack of N-terminus insertion, D194 rotamer, and concurrent 'kink' in the backbone on the surrounding residues. We shall analyze the structural and dynamical consequences caused by this modification in FVII.

The unrestrained simulated structure (in aqueous environment) drifts away from the X-ray crystallographic structure as measured by the decrease of the calculated cross-correlation coefficients. This is probably to be expected because of the temperature differences between computation and experiment, as well as crystal packing.

The RMSF was calculated for each residue of the protein-aligned trajectory to reveal areas of the FVII molecule exhibiting large fluctuations. In the SP domain, flexible areas encompass solvent exposed surface loops (activation linker region, 60-loop, calcium loop, 99-loop, act loop 1, 170-loop, act loop 2, and act loop 3), which is expected considering the domain topology.

In the model FVII, corresponding mass-weighted density volumetric maps was computed for each frame of the entire simulation trajectory, and combined using averaging. This this was done both for the missing parts of activations loop 1, 2, and 3, and all three in combination. The results indicate that said loops are indeed very flexible and no clear conformity between the electro density maps and the simulated density tracings could be

found.

Acknowledgement

The authors thank A. Amstrup and I. Svendsen for excellent technical assistance and H. Østergaard for scientific discussions.

References

1. Banner, D.W., A. D'Arcy, C. Chène, F.K. Winkler, A. Guha, et al. 1996. The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. Nature. 380: 41–46.
2. Bode, W., and R. Huber. 1976. Induction of the bovine trypsinogen—trypsin transition by peptides sequentially similar to the N-terminus of trypsin. FEBS Lett. 68: 231–236.
3. Huber, R., and W. Bode. 1978. Structural basis of the activation and action of trypsin. Acc. Chem. Res. 266: 114–122.
4. Tompa, P. 2002. Intrinsically unstructured proteins. Trends Biochem. Sci. 27: 527–33.
5. James, L.C., and D.S. Tawfik. 2003. Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. Trends Biochem. Sci. 28: 361–8.
6. Higashi, S., N. Matsumoto, and S. Iwanaga. 1996. Molecular Mechanism of Tissue Factor-mediated Acceleration of Factor VIIa Activity. J. Biol. Chem. 271: 26569–26574.
7. Dickinson, C.D., C.R. Kelly, and W. Ruf. 1996. Identification of surface residues mediating tissue factor binding and catalytic function of the serine protease factor VIIa. Proc. Natl. Acad. Sci. U. S. A. 93: 14379–84.
8. Eigenbrot, C., D. Kirchhofer, and M. Dennis. 2001. The factor VII zymogen structure reveals reregistration of β strands during activation. Structure. 9: 627–636.
9. Olsen, O.H., P.F. Nielsen, and E. Persson. 2004. Prevention of beta strand movement into a zymogen-like position does not confer higher activity to coagulation factor VIIa. Biochemistry. 43: 14096–103.
10. Perera, L., and L.G. Pedersen. 2005. A reconsideration of the evidence for structural reorganization in FVII zymogen. J. Thromb. Haemost. 3: 1543–5.
11. Rand, K.D., T.J.D. Jørgensen, O.H. Olsen, E. Persson, O.N. Jensen, et al. 2006. Allosteric activation of coagulation factor VIIa visualized by hydrogen exchange. J. Biol. Chem. 281: 23018–24.
12. Nemerson, Y., and D. Repke. 1985. Tissue factor accelerates the activation of coagulation factor VII: the role of a bifunctional coagulation cofactor. Thromb. Res. 40: 351–358.

13. Svensson, L.A., L. Thim, O.H. Olsen, and E.M. Nicolaisen. 2013. Evaluation of the metal binding sites in a recombinant coagulation factor VIII identifies two sites with unique metal binding properties. Biol. Chem. 394: 761–5.
14. Mammen, C.B., T. Ursby, Y. Cerenius, M. Thunnissen, J. Als-Nielsen, et al. 2002. Design of a 5-station macromolecular crystallography beamline at MAX-lab. Acta Phys. Pol. A. 101: 595–602.
15. Murshudov, G.N., A.A. Vagin, and E.J. Dodson. 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D. Biol. Crystallogr. 53: 240–55.
16. Adams, P.D., P. V Afonine, G. Bunkóczi, V.B. Chen, N. Echols, et al. 2011. The Phenix software for automated determination of macromolecular structures. Methods. 55: 94–106.
17. Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D. Biol. Crystallogr. 60: 2126–32.
18. The Uniprot Consortium. 2013. Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res. 41: D43–7.
19. Sievers, F., A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7: 539.
20. Walter, J., W. Steigemann, T.P. Singh, H. Bartunik, W. Bode, et al. 1982. On the disordered activation domain in trypsinogen: chemical labelling and low-temperature crystallography. Acta Crystallogr. Sect. B. : 1462–1472.
21. Martyna, G.J., D.J. Tobias, and M.L. Klein. 1994. Constant pressure molecular dynamics algorithms. J. Chem. Phys. 101: 4177–4189.
22. Feller, S.E., Y.H. Zhang, R.W. Pastor, and B.R. Brooks. 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. 103: 4613–4621.
23. Allen, M.P., and D.J. Tildesley. 1989. Computer simulation of liquids. Reprint ed. Oxford University Press.
24. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98: 10089–10092.

25. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, et al. 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103: 8577–8592.
26. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al. 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26: 1781–1802.
27. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 102: 3586–3616.
28. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79: 926–935.
29. Ryckaert, J., G. Ciccotti, and H. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23: 327–341.
30. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14: 33–38.
31. Yang, L., E. Eyal, C. Chennubhotla, J. Jee, A.M. Gronenborn, et al. 2007. Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. Structure. 15: 741–749.
32. Trabuco, L.G., E. Villa, K. Mitra, J. Frank, and K. Schulten. 2008. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure. 16: 673–683.
33. Sichler, K., D.W. Banner, A. D'Arcy, K.-P. Hopfner, R. Huber, et al. 2002. Crystal Structures of Uninhibited Factor VIIa Link its Cofactor and Substrate-assisted Activation to Specific Interactions. J. Mol. Biol. 322: 591–603.
34. Papagrigoriou, E., P.A. McEwan, P.N. Walsh, and J. Emsley. 2006. Crystal structure of the factor XI zymogen reveals a pathway for transactivation. Nat. Struct. Mol. Biol. 13: 557–8.
35. Megyeri, M., V. Harmat, B. Major, Á. Végh, J. Balczer, et al. 2013. Quantitative characterization of the activation steps of mannan-binding lectin (MBL)-associated serine proteases (MASPs) points to the central role of MASP-1 in the initiation of the complement lectin pathway. J. Biol. Chem. 288: 8922–34.

36. Cuesta-Seijo, J.A., and S. Garcia-Granda. 2002. Trypsin as a model for high resolution x-ray diffraction in proteins. Bol. R. Soc. Hist. Nat. Sec. Geol. 97: 123–129.
37. Frishman, D., and P. Argos. 1995. Knowledge-Based Protein Secondary Structure Assignment. PROTEINS Struct. Funct. Genet. : 566–579.

Figure- and table texts

TABLE 1: Statistics on Diffraction Data and Structure Refinement of the FVII zymogen

TABLE 2: Published X-ray crystallographic structures of zymogens with the chymotrypsin fold. [a] indicates zymogen with enzyme-induced structural features (and [b] vice versa).

TABLE 3: Overview of the simulated systems

FIGURE 1: (**A**) X-ray crystallographic structure of the FVII zymogen shown in wire representation. There are two complexes in the unit cell, TF_1:$FVII_1$ (in green and gray, respectively) and TF_1:$FVII_1$ (, red and blue ,respectively). (**B**) Close-up of the SP domain of the TF_1:$FVII_1$ monomer of the crystal structure in wire representation. The missing loop regions of the 170-loop, activation loops 2+3, and activation loop 1 are encircles and labeled. (**C**) Structural alignment of the FVII crystal structure (gray), the FVIIa from PDB entry code 1dan(1) (orange), and trypsinogen from PDB entry code 2ptn(20).

FIGURE 2: Crystallographic B factors (rescaled by $\frac{3}{8\pi^2}$) *vs.* residue for the FVII zymogen C_α atoms.

FIGURE 3: RMSF *vs.* residue for the simulated FVII model. Errorbars indicating the standard deviation are drawn as dotted lines. Highlighted above the panel are named regions of the SP domain of the FVII zymogen (brown bars).

FIGURE 4: Cross-correlation coefficient between the X-ray crystallographic electron density map and each frame of the FVII homology model simulation (SP domain only).

FIGURE 5: Comparison of residual density around the missing loop segments in the feature-enhanced crystal density map (FEM) and calculated volumetric maps of the density tracing based in the simulated FVII model. (**A**) Residual density around activation loops 2 and 3 in the electron density (FEM) maps (*top*, at $\sigma = 2.5$; *bottom*, at $\sigma = 1.0$). The protein Cα's are traced out in pink wire and density maps are shown as blue mesh (proteins are oriented similarly). The white arrows indicate density which is likely due to either activation loops 1, 2, 3, or a combination. (**B**) Calculated density tracing of the activation loop 1 isolated (*top left*, in pink). Calculated density tracing of the activation loop 2 isolated (*top right*, in blue). Calculated density tracing of the activation loop 3 isolated (*bottom left*, in red). Calculated density tracing of the activation loops 1+2+3 combined (*bottom right*, in black). Molecular structures are shown in wire representation with the SP domain of the FVII model in gray and TF in green. Shown in sticks are the catalytic triad residues (D242, H193, and S344) as well as W364. The loop segments corresponding to the missing parts from the crystal structure have been colored (act loop 1 in pink; act loop 2 in blue; act loop 3 in red)

FIGURE 6: Scatter-plot showing the structural integrity of the catalytic triad residues in FVII as determined by interatomic C_β-C_β distance of H193 to S344 and H193 to D242 for FVII. The blue arrow indicates the drift (in time.)

FIGURE S1: RMSF *vs.* residue for the reference simulations of trypsin (*top*, PDB entry code 1j8a(36)) and trypsinogen (*bottom*, PDB entry code 2ptn(20)).

FIGURE S2: Scatter-plot showing the structural integrity of the catalytic triad residues for the reference simulations of trypsin (PDB entry code 1j8a(36)) and trypsinogen (PDB entry code 2ptn(20)) as determined by interatomic C_β-C_β distance of H57{c} to S195{c} and H57{c} to D102{c}.

FIGURE S3: PROPKA(v.3) pKa value estimates for the active site histidine (residue H193) during the FVII simulation.

FIGURE S4: Assignment (by STRIDE(37)) of secondary structure during FVII simulation. Chain F is FVII zymogen (residues 1-406), chain T is tissue factor. The one-letter secondary structure codes are: H, α-helix; G, 3_{10}-helix; I, PI-helix; E, extended conformation (sheet); B, isolated bridge; T, turn; C, random coil (none of the previous). (10 frames = 1 ns simulation)

Tables

Collected Crystal Data	Data
Resolution (Å)	∞ – 2.700
Space group	$P4_3$
Unit cell parameters	a = b = 150.530 Å, c = 92.060 Å
	$\alpha = \beta = \gamma = 90^o$
No. of unique reflections	56562
Completeness (%)	99.85
Refinement	
Resolution (Å)	29,521 – 2.700
R_{free} (%)	27.68
RMSDs	
Bond lengths (Å)	0.019
Bond angles (deg)	0.788

TABLE 1

Serine endopeptidase zymogen	PDB entry code	Resolution [Å]	Uniprot entry code for S1 peptidase region	Comments
α-Lytic protease	4pro	2.40		Complex with pro-region
Chymotrypsinogen	2cga	1.80	sp|P17538	Native
	1pyt	2.35		+proproteinase E
Complement protease C1r	1md7,1md8	3.20, 2.80	sp|P00736|464-702	
	1gpz	2.90		Catalytic domain only
Complement protease C1s	4j1y	2.66	sp|P09871|438-680	
Factor VII	1jbu	2.00	sp|P08709|213-452	Re-registration has been refuted
	(this study)	2.7		AL1,2,3, 170-loop not resolved
Factor XI	2f83	2.87	sp|P03951|388-623	AL1,2,3 not resolved
MASP-1	4igd	2.50	sp|P48740|449-696	
MASP-2	1zjk	2.18	sp|O00187|445-684	
Plasminogen	1qrz	2.00	sp|P00747|581-808	Proenzyme domain only
Prethrombin-1	3nxp	2.20	sp|P00734|364-618	
Prethrombin-2	1hag	2.00	sp|P00734|364-618	+hirugen
	1mkx (K)	2.20		+α-thrombin (H)
	3sqe, 3sqh	1.90, 2.20		S195A, *open* form
	1nu7[a]	2.20		+Staphylocoagulase
Profactor B	3hrz (D), 3hs0 (D,I)	2.20	sp|P81187|477-754	+cobra venom factor
Profactor D	1fdp	2.10	sp|P00746|26-253	
Progranzyme K	1mza,1mzd	2.23, 2.90	sp|P49863|27-259	

Prokallikrein 6	1gvl	1.80	sp|Q92876|22-242	
Proproteinase E	1fon 1pyt	1.70 2.35		Truncated +chymotrypsinogen
Tissue-type Plasminogen Activator	1bda	3.35	sp|P00750|311-561	Single chain +synthetic inhibitor
Trypsinogen	2tgt 1tgn, 2ptn, +…	1.70 1.65, 1.55	tr|Q5NV56	
Urokinase-type Plasminogen Activator	4dw2[b]	2.97		uPA in complex with zymogenizing inhibitory antibody Fab

TABLE 2

Serine (pro)protease	Bound cofactor(s)	PDB entry code source	Simulated time [ns]	No. of atoms
Human FVII	Tissue factor, Ca^{2+}	X-tal structure (this study); 2ptn	100	~93,000
Human FVII	Ca^{2+}	X-tal structure (this study); 2ptn	*pending*	~87,000
Bovine trypsin	Ca^{2+}	1j8a	200	~16,000
Bovine trypsinogen	Ca^{2+}	2ptn	150	~15,000

TABLE 3

Figures

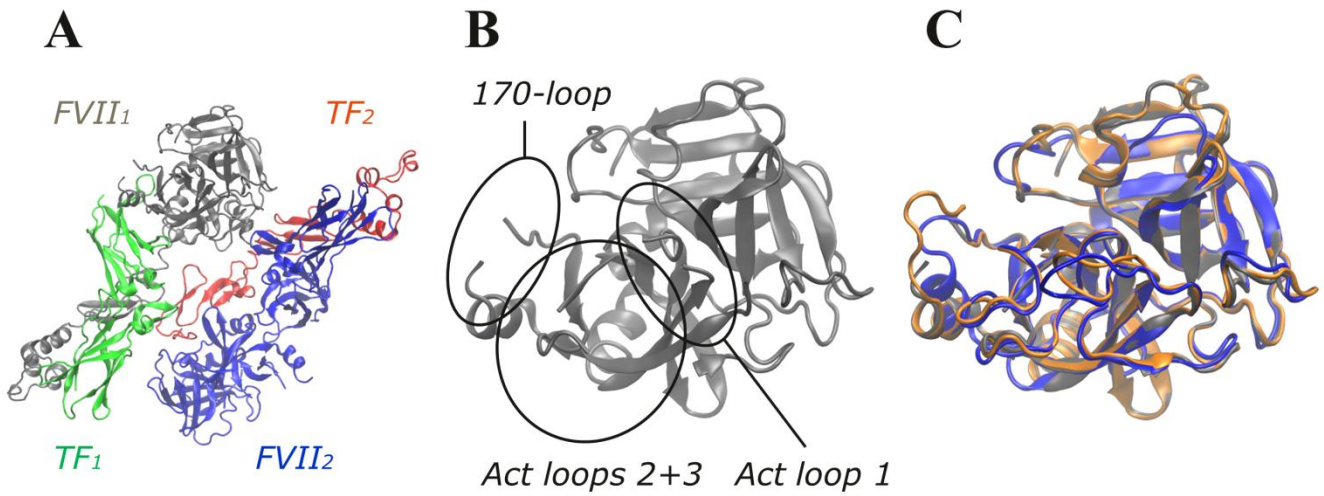


FIGURE 1

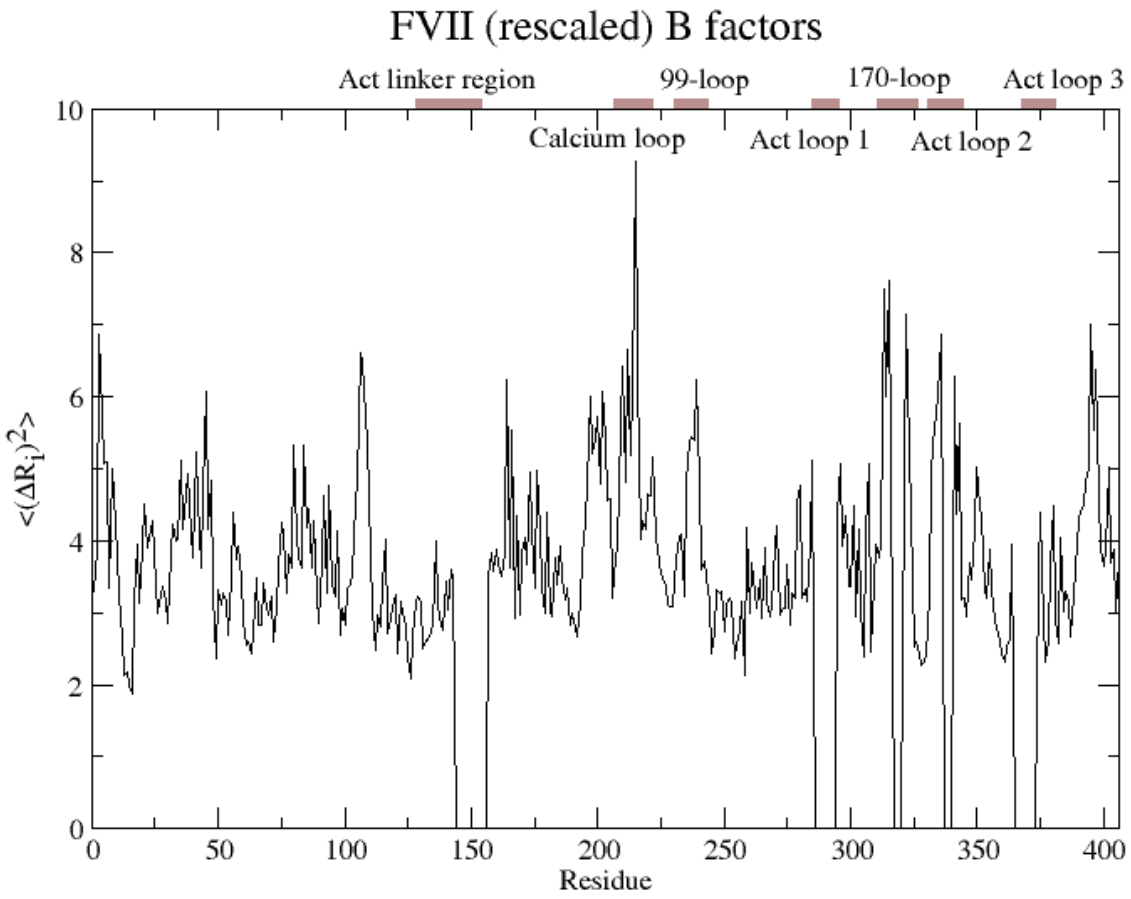

FVII (rescaled) B factors
Act linker region
99-loop
170-loop
Act loop 3
Calcium loop
Act loop 1
Act loop 2
<(ΔR_i)^2>
Residue
0
2
4
6
8
10
0
50
100
150
200
250
300
350
400

FIGURE 2

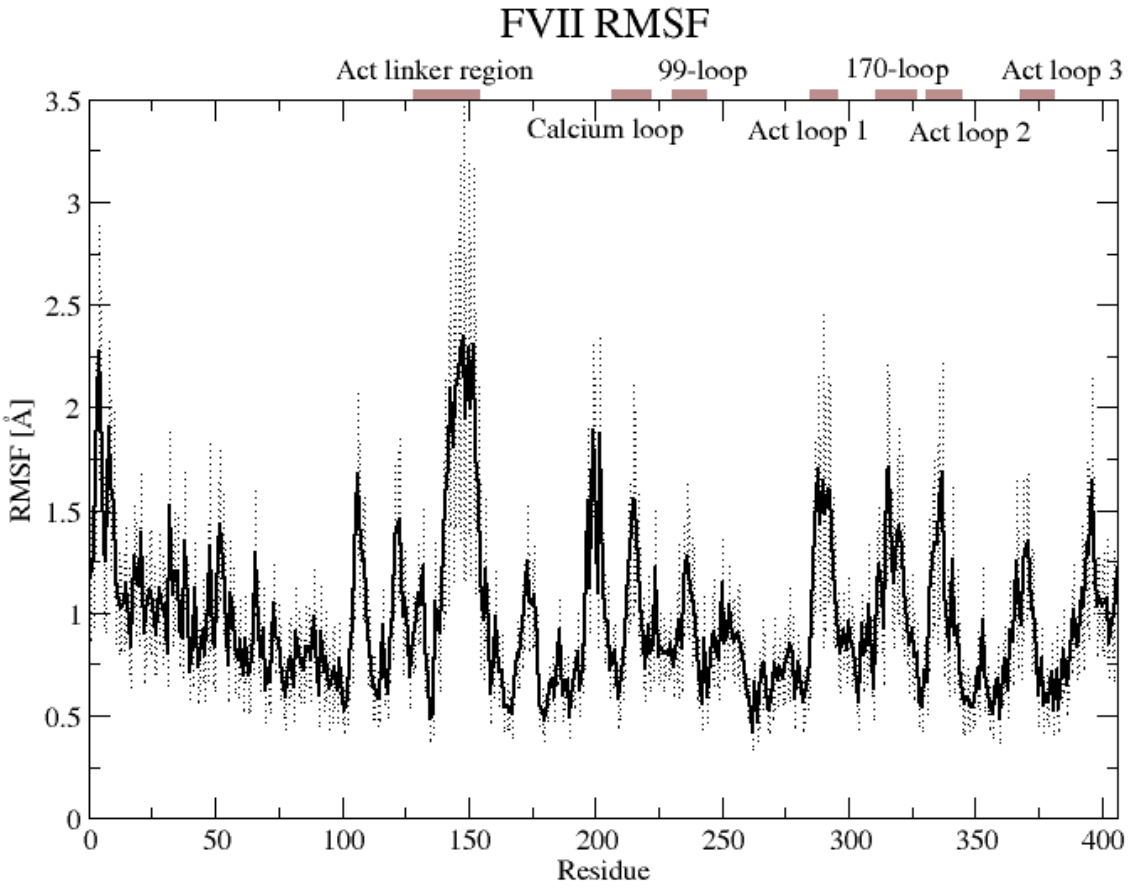

FVII RMSF
Act linker region
99-loop
170-loop
Act loop 3
Calcium loop
Act loop 1
Act loop 2
RMSF [Å]
Residue
0
0.5
1
1.5
2
2.5
3
3.5
0
50
100
150
200
250
300
350
400

FIGURE 3

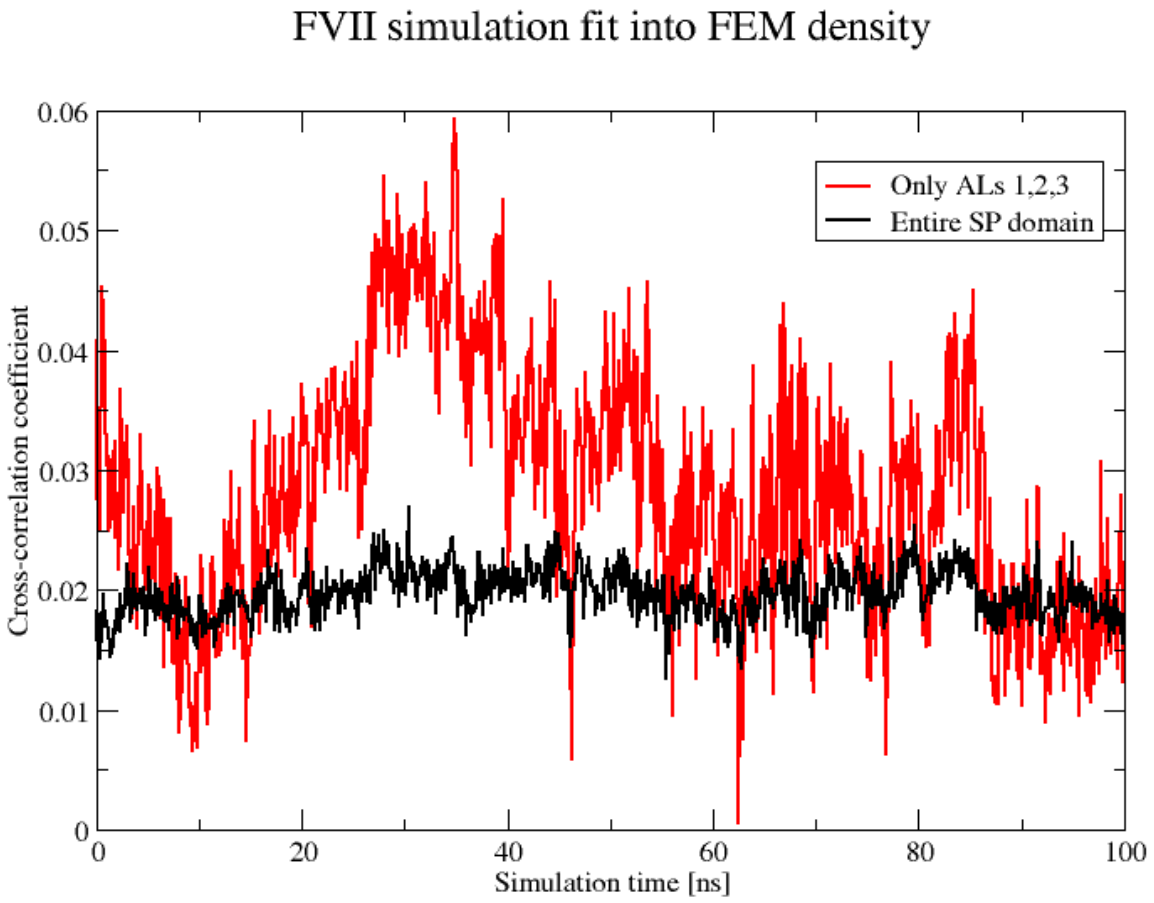

FVII simulation fit into FEM density
Only ALs 1,2,3
Entire SP domain
Cross-correlation coefficient
Simulation time [ns]
0.06
0.05
0.04
0.03
0.02
0.01
0
0
20
40
60
80
100

FIGURE 4

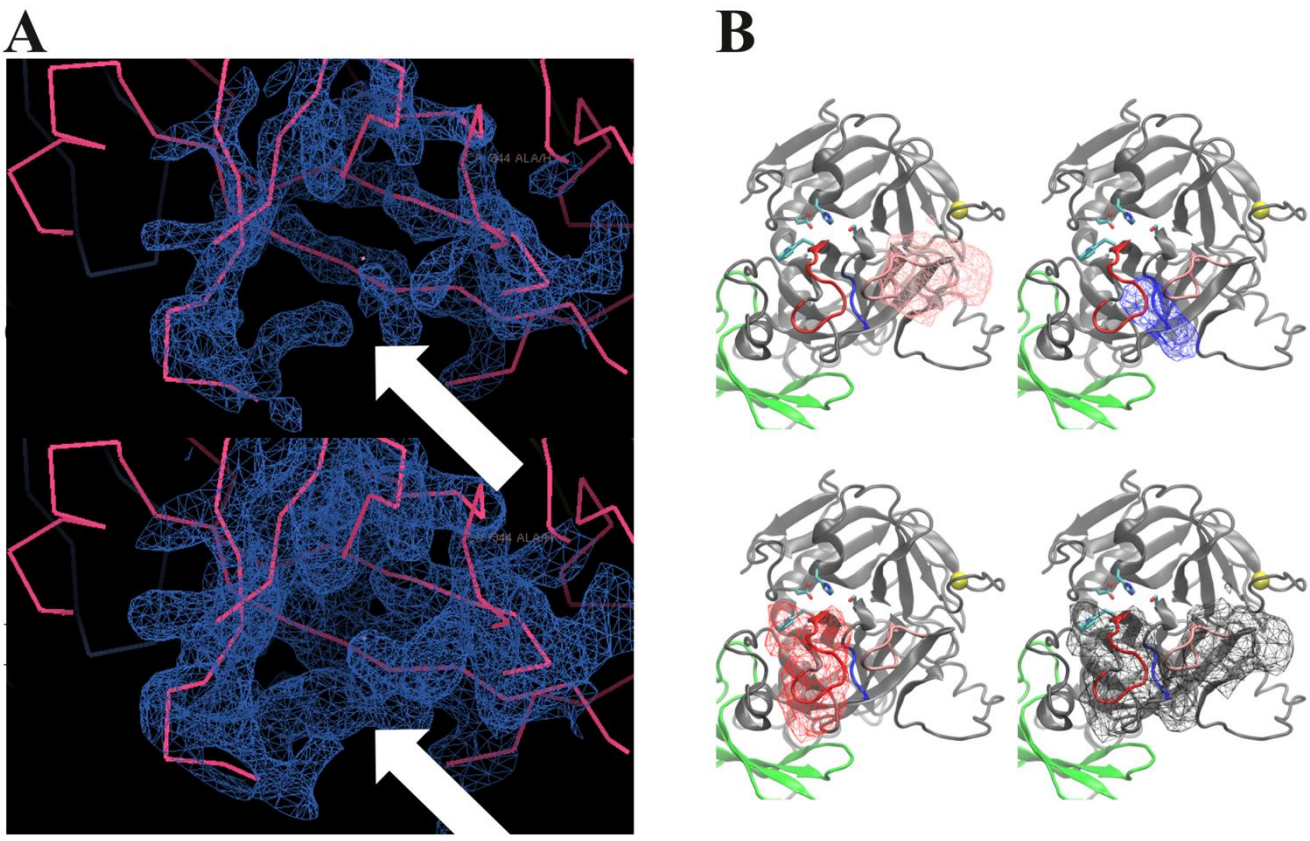


FIGURE 5

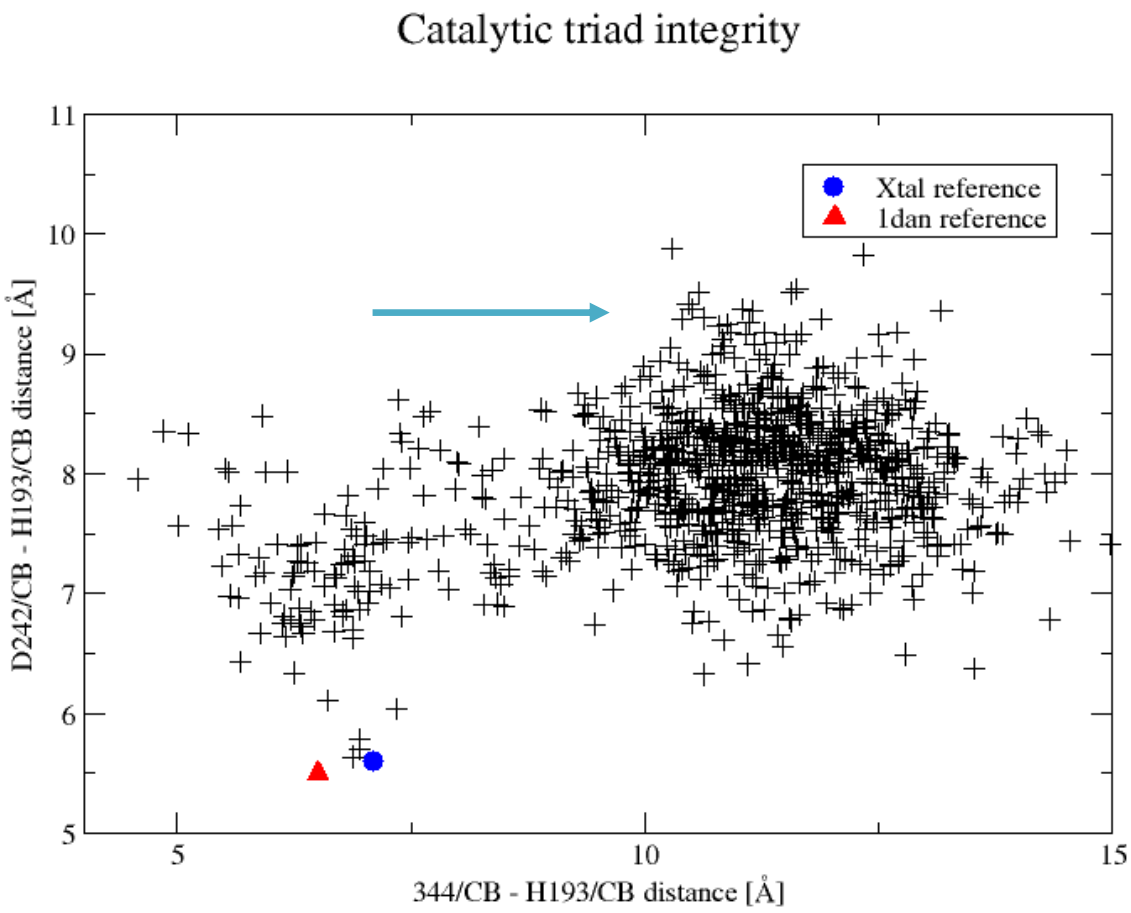

Catalytic triad integrity
Xtal reference
1dan reference
D242/CB - H193/CB distance [Å]
344/CB - H193/CB distance [Å]
11
10
9
8
7
6
5
5
10
15

FIGURE 6

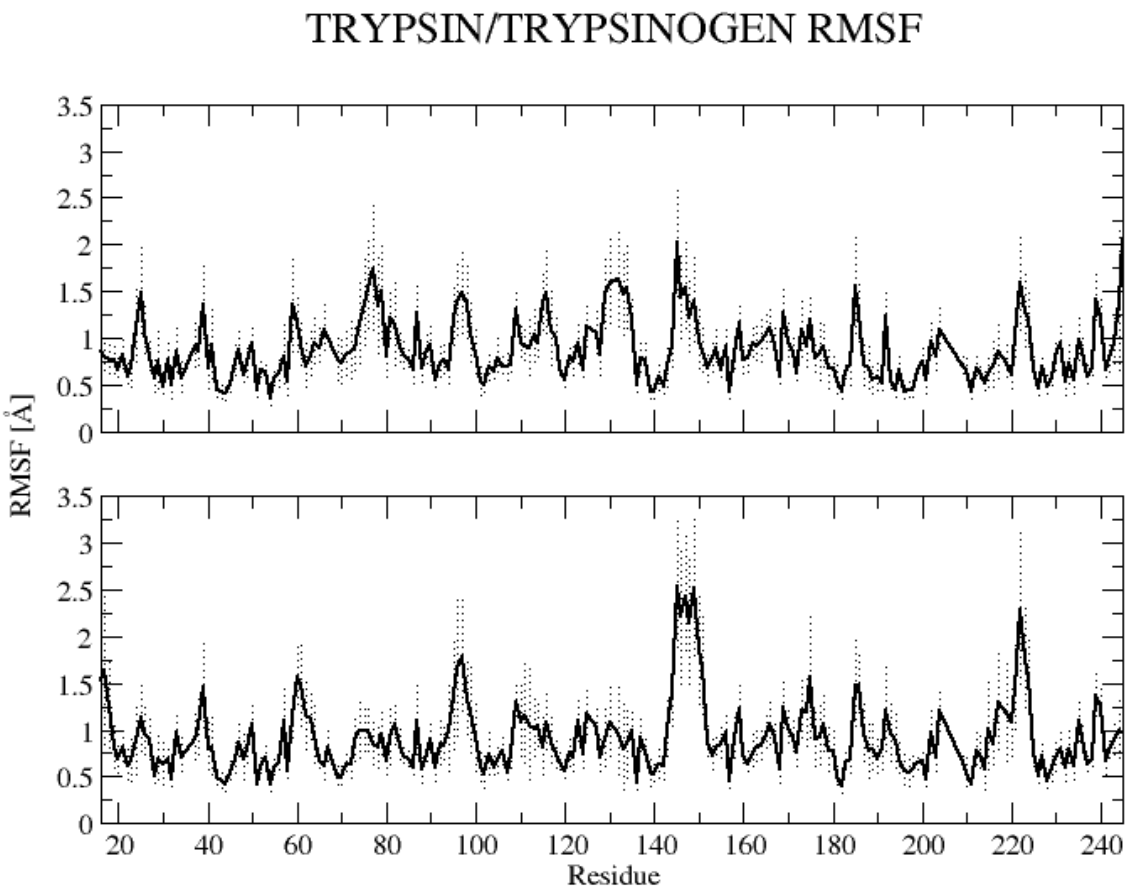

TRYPSIN/TRYPSINOGEN RMSF
RMSF [Å]
Residue
0
0.5
1
1.5
2
2.5
3
3.5
20
40
60
80
100
120
140
160
180
200
220
240

FIGURE S1

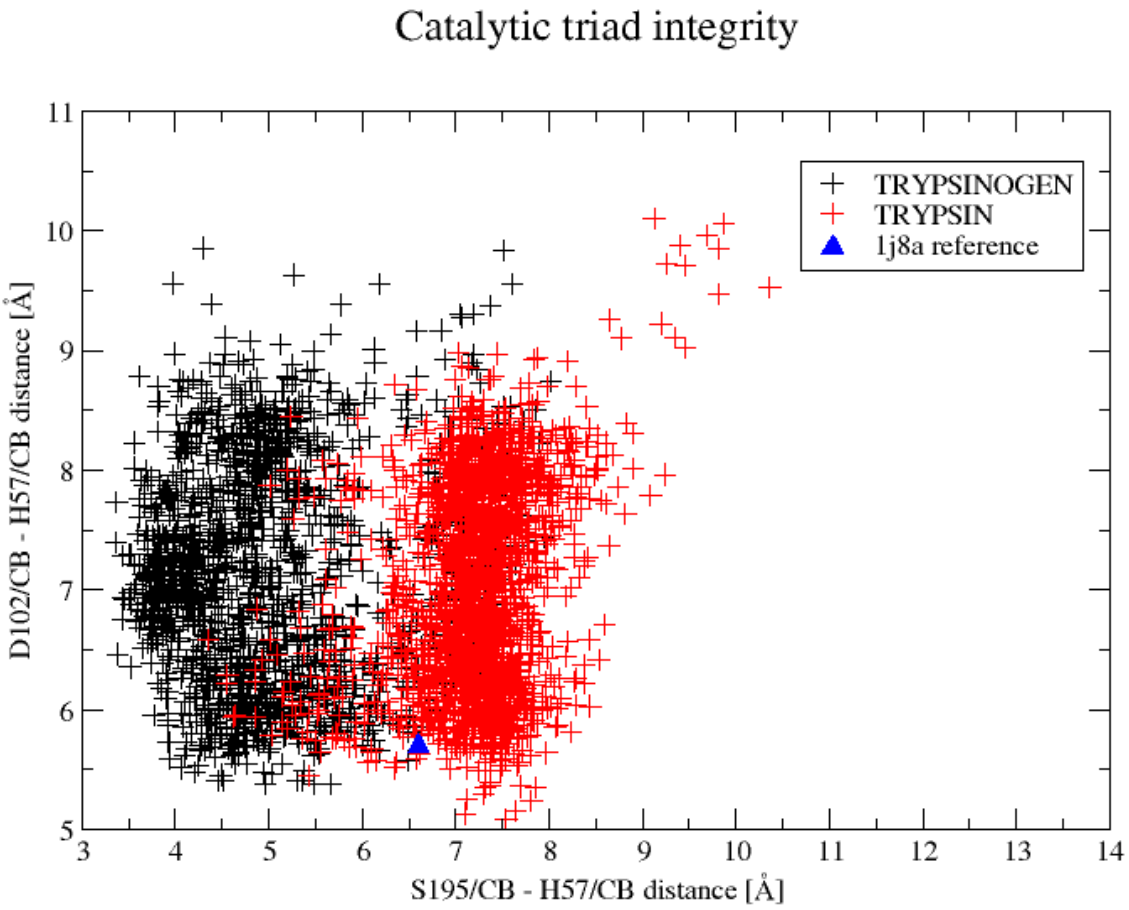

Catalytic triad integrity
TRYPSINOGEN
TRYPSIN
1j8a reference
D102/CB - H57/CB distance [Å]
S195/CB - H57/CB distance [Å]
11
10
9
8
7
6
5
3
4
5
6
7
8
9
10
11
12
13
14

FIGURE S2

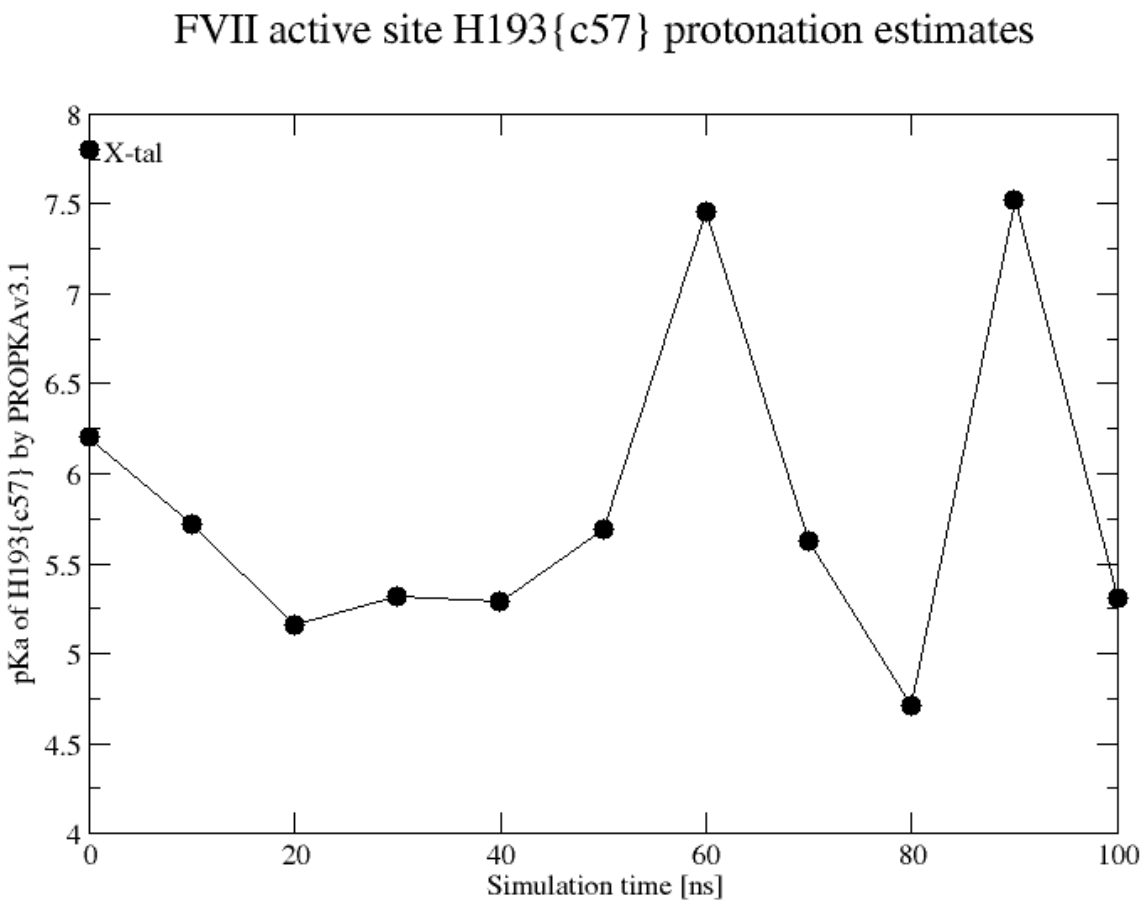

FVII active site H193{c57} protonation estimates
X-tal
pKa of H193{c57} by PROPKAv3.1
Simulation time [ns]
8
7.5
7
6.5
6
5.5
5
4.5
4
0
20
40
60
80
100

FIGURE S3

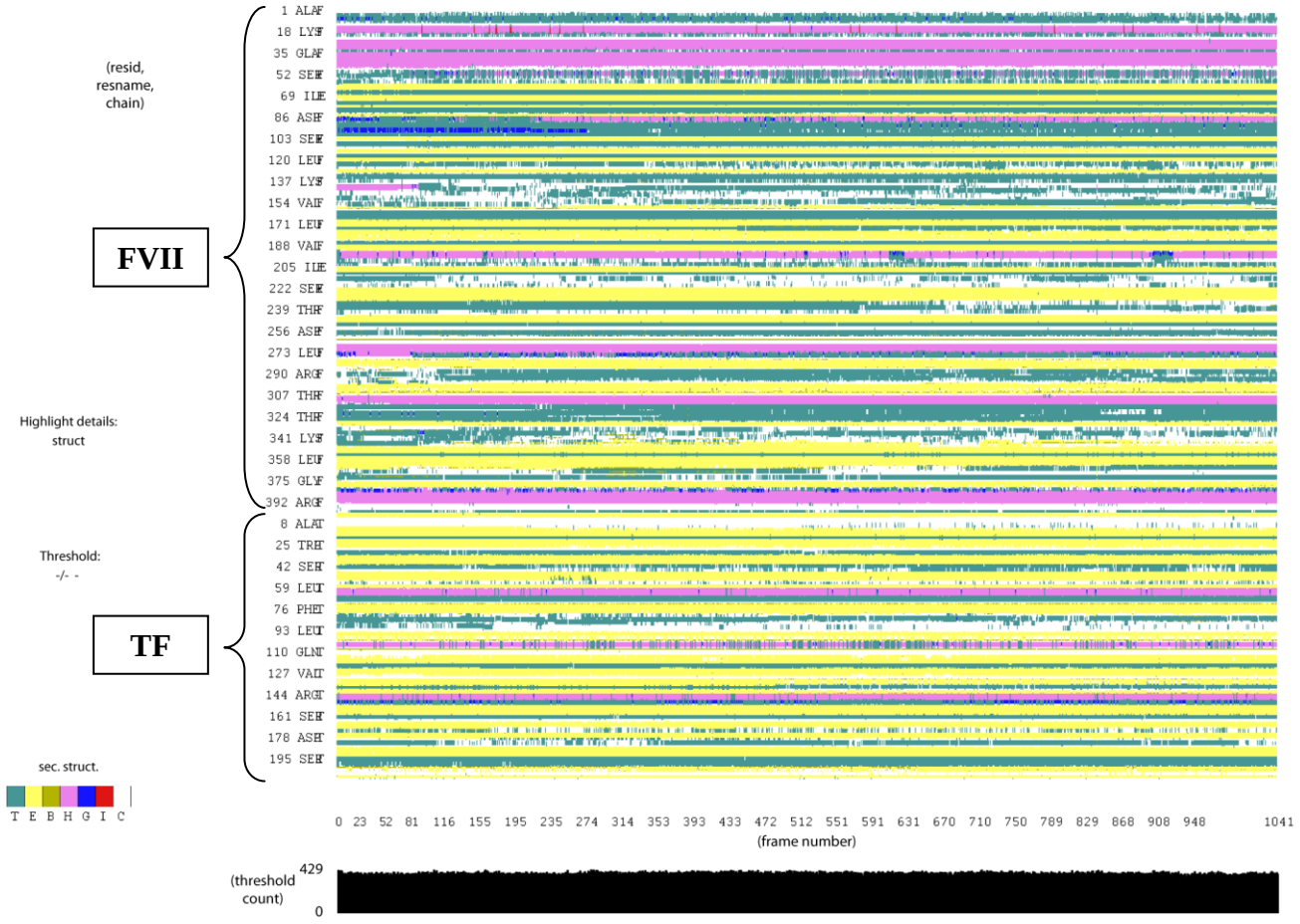

(resid, resname, chain)
FVII
Highlight details: struct
Threshold: -/- -
TF
sec. struct.
T E B H G I C
1 ALAF
18 LYSF
35 GLAF
52 SERF
69 ILEF
86 ASNF
103 SERF
120 LEUF
137 LYSF
154 VALF
171 LEUF
188 VALF
205 ILEF
222 SERF
239 THRF
256 ASNF
273 LEUF
290 ARGF
307 THRF
324 THRF
341 LYSF
358 LEUF
375 GLYF
392 ARGF
8 ALAT
25 THRT
42 SERT
59 LEUT
76 PHET
93 LEUT
110 GLNT
127 VALT
144 ARGT
161 SERT
178 ASNT
195 SERT
0 23 52 81 116 155 195 235 274 314 353 393 433 472 512 551 591 631 670 710 750 789 829 868 908 948 1041
(frame number)
(threshold count)
429
0

FIGURE S4

3.2 Allosteric activation of FVIIa

3.2.1 Prelude

Manuscript 2 entitled “Tissue factor activates allosteric networks in factor VIIa through structural and dynamical changes” by Jesper J. Madsen, Egon Persson and Ole H. Olsen describes the application of MD simulations to elaborate how tissue factor activates coagulation factor VIIa through an allosteric mechanism. Our results elegantly complements previous MD studies of the TF:FVIIa complex(Chang Jun Lee, Wu, Bartolotti, & Pedersen, 2012; Y Z Ohkubo, Morrissey, & Tajkhorshid, 2010) which have focused solely on the enzyme-substrate colocalization, but not described how the protease domain of FVIIa is stimulated by TF.

The work presented in the subsequent notes follows up on the intriguing observation that FVIIa with the 170-loop grafted from trypsin has increased catalytic activity in the absence of TF.(Soejima et al., 2001a) In the following, said FVIIa variant is named FVIIa-311. Several point mutations have been made on top of this FVIIa-311 scaffold to investigate the cause of the increased activity. Particularly promising is the observation modifying the tyrosine residue which harboured by the trypsin 170-loop (to either F for FVIIa-860 or S for FVIIa-313) seems to abrogate the increased activity of these variants (without TF), which, in addition, are not stimulated significantly by TF either. The performed MD simulations of variants FVIIa-311, FVIIa-313, and FVIIa-860 in the absence of TF indicate that the stabilizing effect of the grafted 170-loop is indeed only observed for FVIIa-311, in excellent agreement with the experimental results.

Jointly, manuscripts 2 and the subsequent notes suggest the hypothetical common mechanism of stimulation for TF-activated FVIIa and FVIIa-311 that involves the stabilization of residues 364-368 and a tryptophan-gate via W364.

3.2.2 Manuscript 2: Madsen JM, *et al*, submitted

Tissue factor activates allosteric networks in factor VIIa through structural and dynamical changes

Jesper J. Madsen[1,2], Egon Persson[3], Ole H. Olsen[1]

[1] Haemophilia Biochemistry and [3] Haemophilia Biology, Novo Nordisk A/S, Måløv, Denmark.
[2] DTU Chemistry, Technical University of Denmark, Kgs. Lyngby, Denmark.

Correspondence: Ole H. Olsen, Haemophilia Biochemistry, Novo Nordisk A/S, Novo Nordisk Park, F5.1.29, DK-2760 Måløv, Denmark.
Tel.: +45 3075 4511
E-mail: oho@novonordisk.com

Running Title: Tissue factor-induced allostery in factor VIIa

Summary

Background. Tissue factor (TF) is known to promote colocalization of enzyme (factor (F)VIIa) and substrate (FX or FIX) and to stabilise the active conformation of FVIIa. Details on how the interaction with TF induces structural and dynamical changes in the catalytic domain of FVIIa to enhance its efficiency remain elusive.

Objective. To elucidate the activation of allosteric networks in the catalytic domain of the FVIIa protease when bound to TF.

Methods. Long-timescale molecular dynamics simulations of FVIIa, free and in complex with TF, have been executed and analysed by dynamical network analysis.

Results. Allosteric paths of correlated motion from the TF contact point, M306, in FVIIa to the active site triad can be described and quantified. In particular, shortest paths from M306 to S344 and H193 are 16% and 8% longer in free FVIIa as compared with TF:FVIIa. Furthermore, paths from M306 to I153 (N-terminus) and W364, both representing hallmark residues of allostery, are 7% and 37% longer, respectively, in free FVIIa. There is a significantly weaker coupling between the TF contact point and key residues in the catalytic domain of FVIIa which causes the active site triad to disintegrate when TF is not present.

Conclusions. These findings complement the current understanding on how the protease FVIIa is stimulated by TF. In addition to the allosteric paths previously identified by mutagenesis of FVIIa, we demonstrate allosteric networks in the catalytic domain that are activated by TF and help making FVIIa an efficient catalyst of FIX and FX activation.

Keywords: Allosteric Regulation; Coagulation Factor VIIa; fXase; Molecular Dynamics Simulations; Tissue Factor

Introduction

Factor (F)VIIa is the initiating protease in tissue factor (TF)-triggered blood coagulation. Upon trauma in the form of blood vessel injury, circulating FVIIa encounters its cofactor, TF, to assemble a highly potent complex which launches the molecular cascade that eventually leads to a burst of thrombin and ensuing fibrin clot formation. The details of the allosteric mechanism by which TF transforms FVIIa to an efficient catalyst of FIX and FX activation remain elusive and warrant further investigation.

A major breakthrough toward understanding the initiation of coagulation on the molecular level was provided with the solution of the X-ray crystallographic structure of the binary complex between FVIIa and soluble TF[1]. To this date, most of our structural knowledge on TF-induced changes and allostery in FVIIa is rationalized on the basis of the structure of the TF:FVIIa complex or inferred from the comparison of this structure with free FVIIa structures[2]. Within the last decade, accumulating experimental evidence in the form of mutagenesis studies combined with biophysical and functional characterization have laid the foundation for two distinct TF-induced allosteric pathways in FVIIa[2]. Path I consists of the stabilization of the TF-binding helix (residues 307{c165}-312{c170}) that propagates through the 170-loop (residues 313{c107A}-321{c170I}) to activation loop 3 (residue 372{c223}) causing a maturation of substrate-binding subsites S1-S3. In an orthogonal path (path II), the favorable interaction between L305{c163} and F374{c225} induced by TF binding changes the conformation of activation loop 3 (residues 364{c215}-372{c223}) , affecting the activation pocket and the S1 subsite. This in turn encourages N-terminal insertion and the formation of a salt bridge with D343{c194}. It is well noteworthy that there exists variants of FVIIa, e.g. $FVIIa_{DVQ}$[3], known to have the N-terminus protected from chemical modification (presumably inserted into the activation pocket) which are still significantly activated by TF. This observation would suggest that TF-induced allostery is present in both populations (protease-like and zymogen-like) of the FVIIa ensemble, and furthermore that the allosteric signal is accentuated for molecular states of FVIIa where the N-terminus is not properly inserted into the activation pocket. Common to path I and II is that M306{c164} is the main origin of TF-induced effects on the conformation and activity of FVIIa[4].

There is no structure available for native (free and uninhibited) FVIIa. This is unfortunate because inhibitors, covalent or not, invariably stabilize the active conformation of the FVIIa protease domain by inducing structural changes similar to those mediated by TF. In addition, X-ray crystallographic structures are static snapshots and reveal little or nothing about dynamic properties. One tool to extend and add to this picture a dynamical dimension with atomic level detail is provided by computational biophysics and the molecular dynamics (MD) method.

Recently, an intriguing study[5] showed that it is possible to interpret MD simulations using dynamical network analysis[6]. In doing so, the authors demonstrated qualitatively how a subtle redistribution of the dynamical ensemble of thrombin could separate the behavior of the protease domain upon performing two diametrically opposed biological activities regulated by cofactor binding. For FVIIa, in particular one computational study[7] has explained in detail how TF manages to restrain FVIIa in a conformation which enhances catalytic activity by enzyme-substrate alignment and colocalization. Similar results have been reported for the binary complex between FVIIa and TF in the presence of the TF transmembrane domain[8]. However, these investigations do not address the mode of action of the allosteric switch originating from the primary TF contact point, M306{c164} [4][9]. Furthermore, the decisive role of TF on the structure of the FVIIa protease domain, the enforcement of N-terminal insertion, is not well understood. Hence, at present, a comprehensive mechanistic picture of the ways TF stimulate FVIIa is still missing.

Materials and methods

We have adopted an approach where the dynamical analysis of Luthey-Schulten and co-workers[6] is performed in a sliding window to analyze long-timescale molecular dynamics (MD) simulations of the free FVIIa protease (PDB entry code 1kli, ~1.0 μs simulation) and the TF:FVIIa complex (based on PDB entry code 1dan, ~0.45 μs simulation) in order to produce novel rationalizations on the allosteric communication networks in FVIIa without and with TF.

Preparation of systems. The coordinates for the TF:FVIIa complex were built as described by Ohkubo *et al.* [7] and were graciously provided by the authors. Briefly, the complex was modeled on the basis of PDB entry code 1dan with missing loop segments modeled from PDB entry codes 1boy and 2hft for TF and PDB entry codes 1qfk and 2h9e for FVIIa. To facilitate simulation of long-timescale dynamics, the free FVIIa structure was based on PDB entry code 1kli which is a truncated construct with the heavy chain and only the EGF2 domain of the light chain. For all structures, the co-crystallized inhibitor was removed. All bound calcium ions were included in the simulations. The resulting systems

of TF:FVIIa and free FVIIa were solvated using the Solvate in VMD[10] and ionized to a salt concentration of 0.15 M with Na^+ and Cl^- ions.

Simulation details. Conventional molecular dynamics simulations were performed using the NAMD2[11] software, the CHARMM27 forcefield[12] and the TIP3P water model[13]. An integration time step of 1.0 femtosecond (fs) was used for the velocity Verlet algorithm. Simulations were carried out at constant pressure (P = 1 atm), and constant temperature (T = 310 K) controlled by the Nosé-Hoover Langevin piston barostat[14] and the Langevin thermostat (damping coefficient: 5/ps), respectively. Throughout, anisotropic pressure coupling was applied for the barostat using a piston period of 100 fs and piston decay of 50 fs. Long-range electrostatic forces were calculated using the Particle Mesh Ewald method[15] using a grid spacing of approximately 1 Å and a fourth-order spline for interpolation. Electrostatic forces were updated every fourth fs. Van der Waals interactions were cut off at 12 Å in combination with a switching function beginning at 10 Å. Periodic boundary conditions were applied in *x*-, *y*-, and *z*-directions. The potential energy in all prepared systems was initially minimized using 5,000 steps of the conjugated gradient method.

Dynamical network analysis.
The dynamical network analysis was performed using the NetworkView plugin of VMD[10] which utilizes Carma[16] to calculate covariance matrices. The NetworkView plugin creates a graph of the protein structure by the following rules: Each C_α atom is assigned a node and nodes are inter-connected by an edge if and only if they are within 5 Å at least 75 % of the trajectory window (excluding main chain-neighboring C_α atoms). A weight, w_{ij}, is calculated for each edge based on the atomic displacement covariance coefficient (correlation value, c_{ij}) by the formula

$$w_{ij} = -\log(|c_{ij}|),$$

which allows the weights to be interpreted as ‘lengths’ or ‘distances’ because a higher (numerical) correlation gets mapped onto a shorter distance. Following this nomenclature, the distance of an allosteric path, D_{ij}, is simply found by summation,

$$D_{ij} = \sum_{k,l} w_{kl}.$$

Sliding window analysis.

A sliding window approach was applied to sample the long-timescale molecular dynamics trajectories. The optimal path from M306 to residues of interest (I153, H193, S344, W364) was found in a 10-ns window whose starting point was incremented by +2 ns in each step to slide over the entire simulated trajectory. The relative observed frequencies were normalized by the number of occurrences of the most frequent (primary) path during the whole simulation trajectory. The average length of the specified paths was calculated by averaging its length in the individual windows where it was found.

Results and discussion

Our main results, sketched in Figure 1, elucidate how the long-range allosteric signal propagates from the TF contact point (M306{c164}) through the protease domain to the catalytic residues of the active site (Figure 1A) causing the triad to be stabilized (Figure 1B). Of particular interest to the propagation of the allostery are structural changes enabling the transmission of the signal through the complexed FVIIa, and the actual residues within which the signal itself travels via correlated motions and fluctuations. In the language of dynamical network analysis, the former affects the residue-residue contact maps describing allowed connectivities in the protein, while the latter defines the strength of these pair-wise interactions. In the TF:FVIIa complex, the predominant pathway of communication (from the source M306{c164}) to the active site residue targets S344{c195} and H193{c57} is ~47 % and 35 % more frequent than the second-most observed path, respectively. Actual optimal paths traveled by the allosteric signal are M306{c164}-C329{c182}-V376{c227}-T378{c229}-D242{c102}-H193{c57} and M306{c164}-C329{c182}-Y377{c228}-V362{c213}-S344{c195}. Comprehensive statistics for lesser traveled paths (less frequent optimal paths and suboptimal paths) is available from the authors upon request. Interestingly, in the absence of TF, the same shortest source-to-target paths M306{c164}-S344{c195} and M306{c164}-H193{c57} are, on average, 16 % and 8 % (based on the full length trajectory) longer and, hence, result in a significantly weaker coupling for free FVIIa than for TF:FVIIa (similar results have been observed upon performing the analysis on the trajectories from Ohkubo *et al.*[7] of TF:FVIIa bound to a membrane patch). Furthermore, the experimentally determined

pathways of allostery summarized above, path I and II, are 37 % and 7 % longer in free FVIIa when represented by the connection between M306{c164}-W364{c215} and M306{c164}-I153{c16}, respectively. The strongest paths of correlated motions, however, do not align with the mechanistic picture suggested by the experimental data[2]. These findings illustrate how TF can induce an allosteric effect within the FVIIa protease domain and, furthermore, suggest through which residues the allosteric signal is most likely to travel on its way from the primary enzyme-cofactor contact point to the catalytic apparatus. As a consequence, the active site is stabilized in the potent, canonical trypsin-like configuration (Figure 1B). While there is significant variance in the distances between the catalytic residues in our dynamical simulation, TF clearly induces structural preservation of the active site triad with gaussian variation around the 1dan reference point. In the simulation of free FVIIa, however, a complete loss of the canonical triad geometry is observed as indicated by the drift of the critical C_β-C_β distances from H193{c57} to S344{c195} and, in particular, to D242{c102} (Figure 1B). In addition, salt bridge between the N-terminal I153{c16} and D343{c194} is significantly destabilized for free FVIIa as gauged by the inter-atomic distance between the γ-carbon (CG) of D343{c194} and the amino group nitrogen (N) of I153{c16}, which is 3.25 ± 0.13 Å for TF:FVIIa and 3.50 ± 0.48 Å for free FVIIa during the course of the simulations. A further structural stabilization is seen in the residues following the ‘E2’ β-strand leading into activation loop 3.

A dynamical picture is emerging from the molecular dynamics simulations of how TF stimulates the productivity of FVIIa in the complex by, in addition to alignment and colocalization of enzyme (FVIIa) and substrate (FIX/FX) as previously described (5)(6), induction of a direct, long-range allosteric effect in the protease domain originating from a contact point at the N-terminal part of the TF-binding helix. This induced effect, in turn, stabilizes the active site catalytic triad as well as encourages N-terminal insertion. The allosteric network illuminated in this study would most likely not be deducible by mutagenesis studies, partly due to the disruptive nature of mutations introduced in the protein core, but also because signal-carrying motions include those of backbone amide and carbonyl groups. Vice versa, the experimentally derived allosteric pathways I and II and inferred mechanisms are not evident from the results of the dynamical network

analysis. However, paths of correlated motions exist between the TF contact point and key structural points representing path I and II which are strengthened in the presence of TF. We are very excited that experimental data supporting the proposed allosteric network are emerging from studies using hydrogen-deuterium exchange mass spectrometry[17](Hongjian Song & Kasper D Rand, personal communication).

Figure text

Figure 1: (A) Primary allosteric paths from the TF-contact point to the active site H193{c57} and S344{c195} residues. TF (green), FVIIa light chain (tan) and heavy chain (cyan) are shown in wire-representation. The catalytic triad residues of the FVIIa protease domain are shown in sticks. Bound Ca^{2+} are shown as spheres (yellow). C_α atoms of the residues of the proposed allosteric paths from M306 to H193 and S344, respectively, are highlighted by spheres (purple) and labeled according to amino acid type and residue numbers (full-length FVIIa numbering). (B) Geometry of the active site triad residues D242{c102}, H193{c57}, and S344{c195} in TF:FVIIa (green dots) and free FVIIa (black dots). Reference values from PDB entry code 1dan[1] are H57:CB/D102:CB = 5.5 Å and H57:CB/S195:CB = 6.5 Å and shown by a red triangle. Top left insert shows the structure of the canonical catalytic triad from PDB entry code 1dan[1] with the catalytic residues shown in sticks with the C_β atoms highlighted as purple spheres.

Addendum

J. J. Madsen designed the research, performed the experiments, analysed results, and wrote the manuscript.

E. Persson and O. H. Olsen designed the research, analysed results, and revised the manuscript.

Acknowledgement

The authors thank Dr. YZ Ohkubo for providing simulation trajectories from their study.

Disclosure of Conflicts of Interest

The authors state they have no conflicts of interest.

References

1 Banner DW, D'Arcy A, Chène C, Winkler FK, Guha A, Konigsberg WH, Nemerson Y, Kirchhofer D. The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature* **1996;** 380: 41–6.

2 Persson E, Olsen OH. Allosteric activation of coagulation factor VIIa. *Front Biosci* **2011;** 16: 3156–63.

3 Persson E, Kjalke M, Olsen OH. Rational design of coagulation factor VIIa variants with substantially increased intrinsic activity. *Proc Natl Acad Sci U S A* **2001;** 98: 13583–8.

4 Persson E, Nielsen LS, Olsen OH. Substitution of aspartic acid for methionine-306 in factor VIIa abolishes the allosteric linkage between the active site and the binding interface with tissue factor. *Biochemistry* **2001;** 40: 3251–6.

5 Gasper PM, Fuglestad B, Komives EA, Markwick PRL, McCammon JA. Allosteric networks in thrombin distinguish procoagulant vs . anticoagulant activities. *Proc Natl Acad Sci U S A* **2012;** 109: 21216–22.

6 Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A* **2009;** 106: 6620–5.

7 Ohkubo YZ, Morrissey JH, Tajkhorshid E. Dynamical view of membrane binding and complex formation of human factor VIIa and tissue factor. *J Thromb Haemost* **2010;** 8: 1044–53.

8 Lee CJ, Wu S, Bartolotti LJ, Pedersen LG. Molecular dynamics simulations of the binary complex of human tissue factor (TF(1-242)) and factor VIIa (TF(1-242) /fVIIa) on a 4:1 POPC/POPS lipid bilayer. *J Thromb Haemost* **2012;** 10: 2402–5.

9 Rand KD, Andersen MD, Olsen OH, Jørgensen TJD, Østergaard H, Jensen ON, Stennicke HR, Persson E. The origins of enhanced activity in factor VIIa analogs and the interplay between key allosteric sites revealed by hydrogen exchange mass spectrometry. *J Biol Chem* **2008;** 283: 13378–87.

10 Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* **1996;** 14: 33–8.

11 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem* **2005;** 26: 1781–802.

12 MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998;** 102: 3586–616.

13 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983;** 79: 926–35.

14 Feller SE, Zhang YH, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* **1995;** 103: 4613–21.

15 Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys* **1995;** 103: 8577–92.

16 Glykos NM. Carma: A Molecular Dynamics Analysis Program. *J Comput Chem* **2006;** 27: 1765–8.

17 Rand KD. Pinpointing changes in higher-order protein structure by hydrogen/deuterium exchange coupled to electron transfer dissociation mass spectrometry. *Int J Mass Spectrom* **2013;** 338: 2–10.

Figures

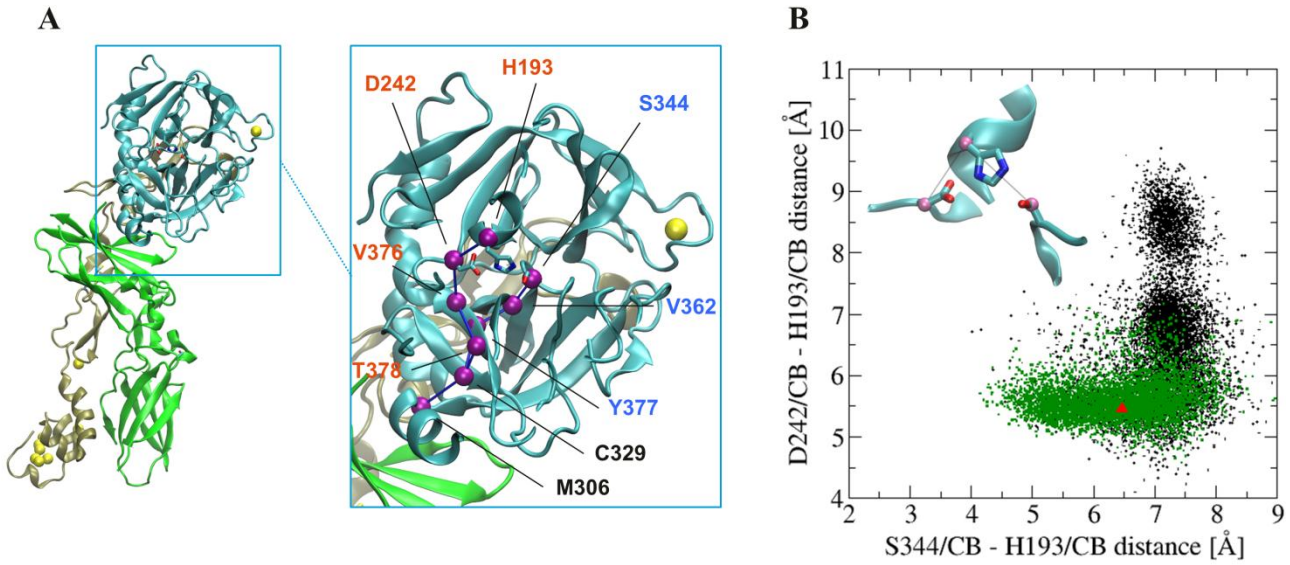


Figure 1

Supporting Information

Tissue factor activates allosteric networks in factor VIIa through structural and dynamical changes

Jesper J. Madsen, Egon Persson, Ole H. Olsen

Materials and Methods

Salt bridge interactions.

Salt bridges with one interacting partner in both FVIIa and TF were detected by applying a geometric distance cut-off equal to 3.2 Å between *any* of the sidechain oxygen atoms of acidic residues and nitrogen atoms of the basic residues. Resulting salt bridges were divided into four groups: those present persistently, present transiently, formed, or broken.

Results and Discussion

In this SI document further analysis and the raw data upon which the distance of each allosteric path is calculated in the free FVIIa and TF:FVIIa simulations. The covariance matrix of the TF:FVIIa simulation supports the experimentally derived fact of M306 being the primary point entry for TF in FVIIa (Fig. S1). The region enclosed the green box, representing M306 in FVIIa and its contact partners (W45 & Y94) in TF, elaborates that the residue-residue correlated motion is very high.

Fig. S2 shows the distance between Cγ atom of D242 and the amino group nitrogen atom of the N-terminus I153, and such indicates the stability of the N-terminally formed salt bridge. Violent fluctuations are evident in the simulations of free FVIIa, wheresas these are effectively truncated by the presence of TF. Furthermore, no significant slip or drift in the interface between FVIIa and TF is observed over the course of the TF:FVIIa simulation. However, not all inter-protein salt bridges are rigidly formed (Figs. S3-S6). The dampening effect of TF on the FVIIa SP domain is gauged by finding the shortest paths from M306 to any other residue of the SP domain (Fig S7). Surpricingly, the ‘dampening’ effects of TF, by which it likely enhances the inter-residue correlated motions in the SP domain of FVIIa is not found for all regions. In particular, the 170-loop seems to be decoupled in the free FVIIa but not in TF:FVIIa.

Tables S1-S5 contain the raw data for the calculated lengths of allosteric paths of interest based on our MD simulations.

Figures and Tables

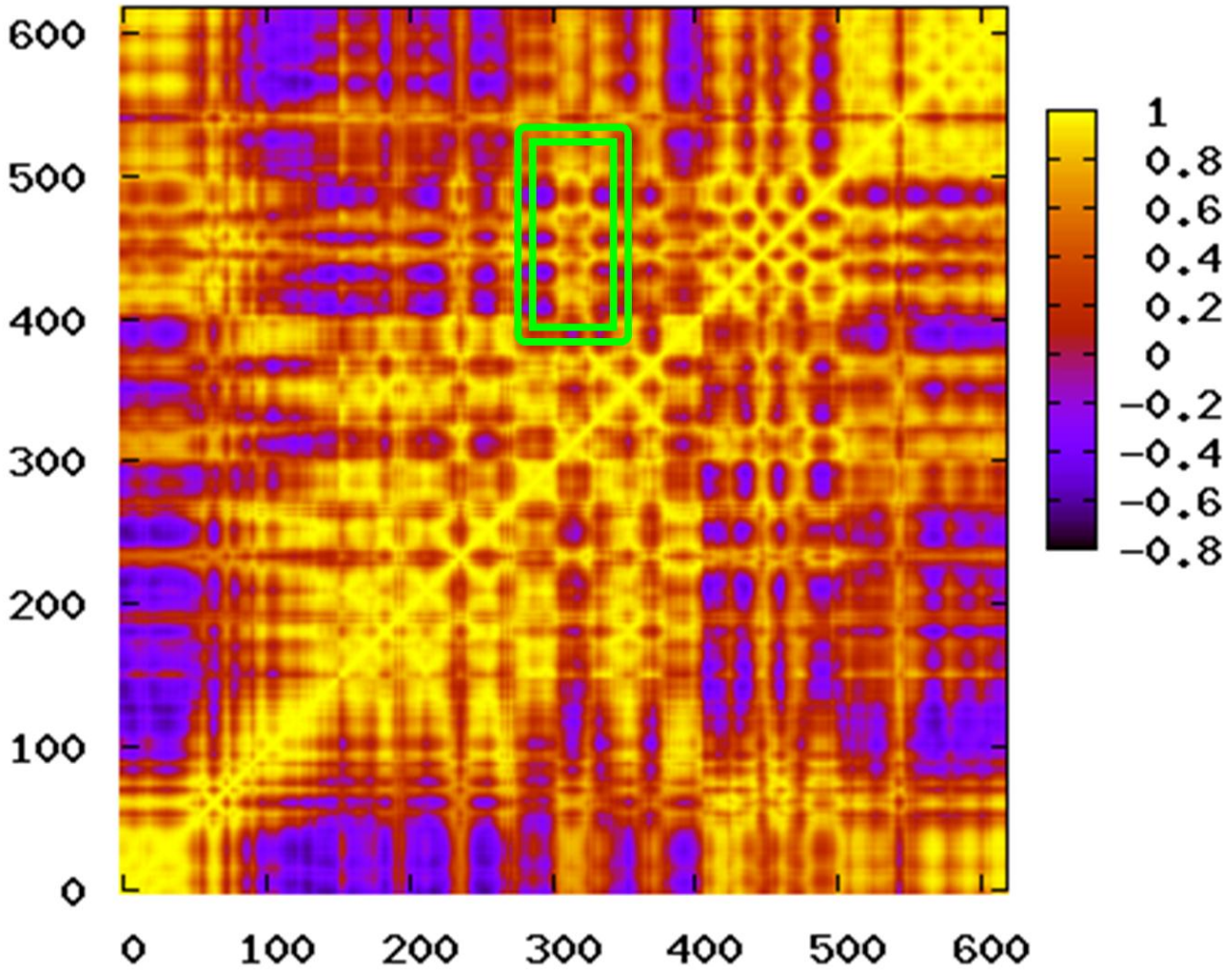


Figure S1: Variance-covariance matrix. Residues 1-406 represent FVIIa and residues 407-619 belong to TF. Highlighted is the region around the primary TF contact in FVIIa, M306{c164}, and its corresponding interaction-partner in TF.

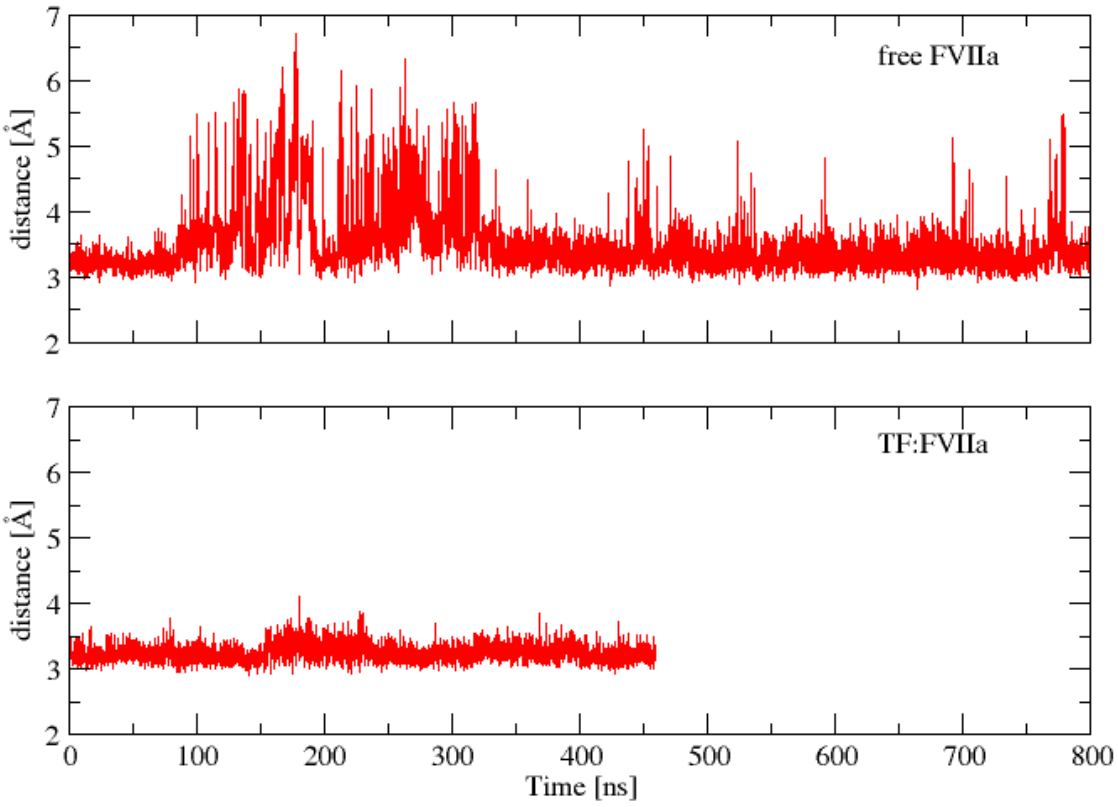


Figure S2: Distance between the C_γ atom of D242 and the amino group nitrogen atom of the N-terminus I153 plotted for the long-timescale simulations of free FVIIa (top) and the binary complex between TF and FVIIa (bottom).

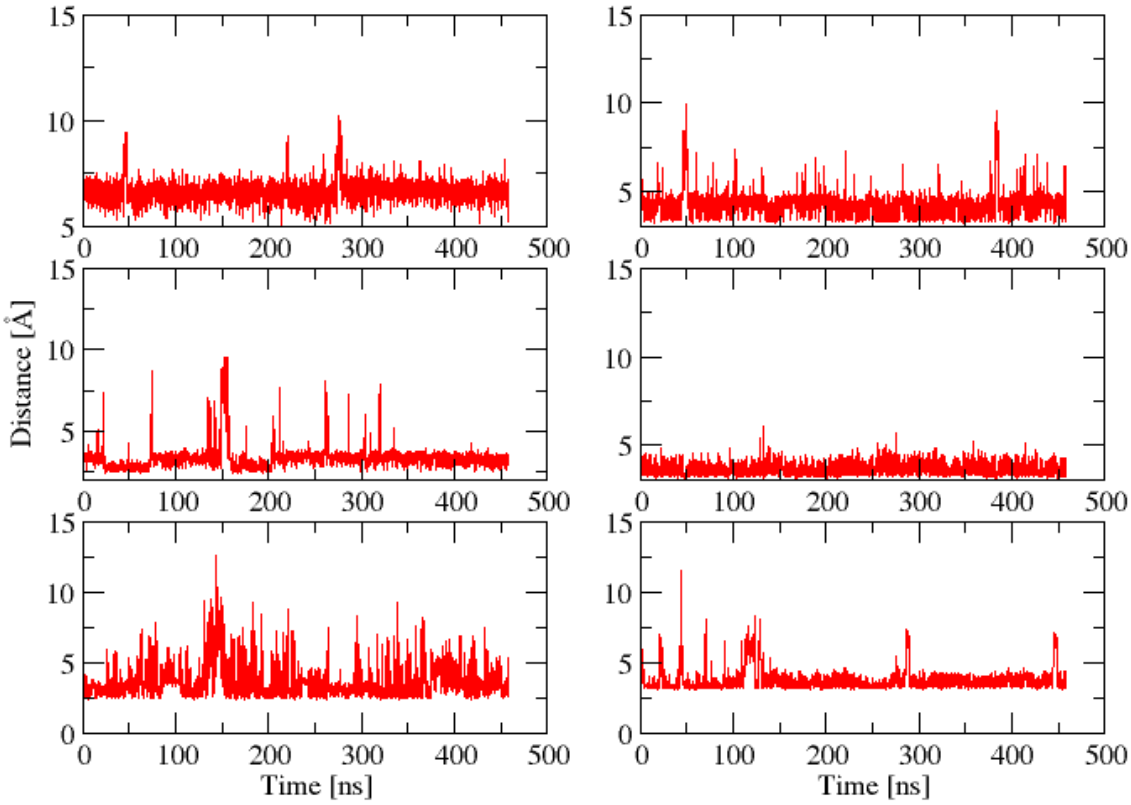


Figure S3: Persistent salt bridges between TF and FVIIa in the simulation of the tenase complex.

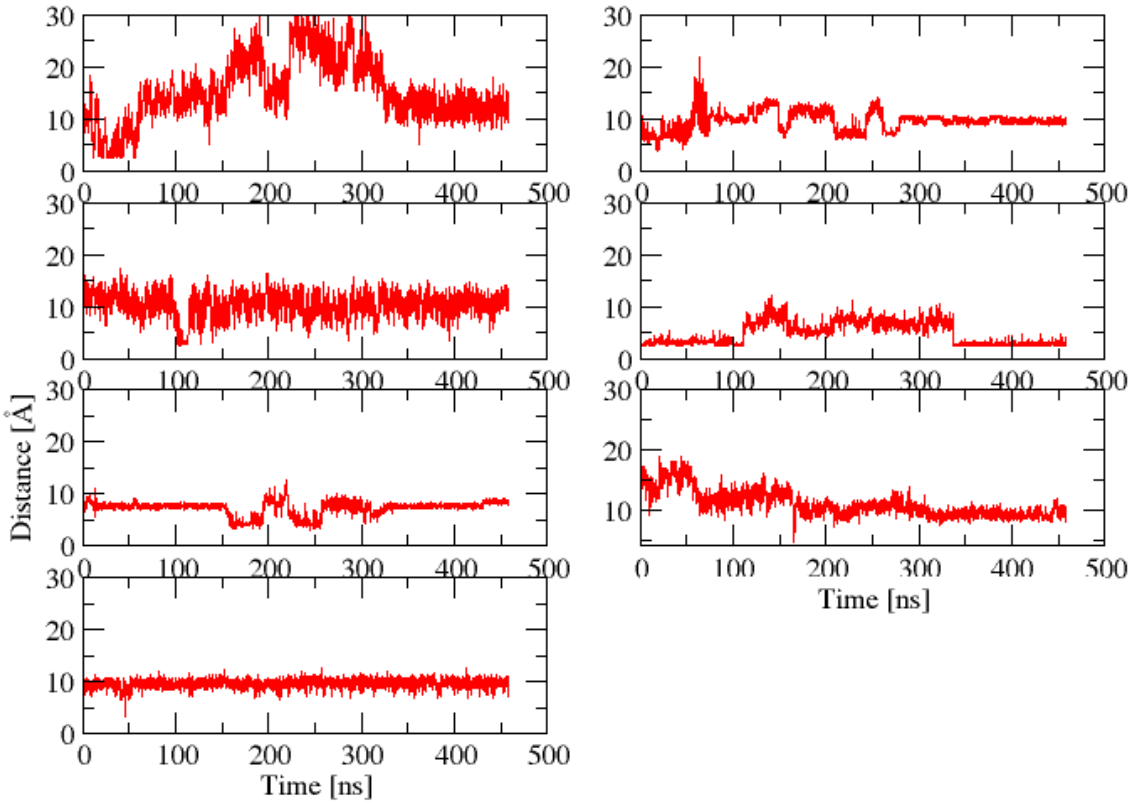


Figure S4: Transient salt bridges between TF and FVIIa in the simulation of the tenase complex.

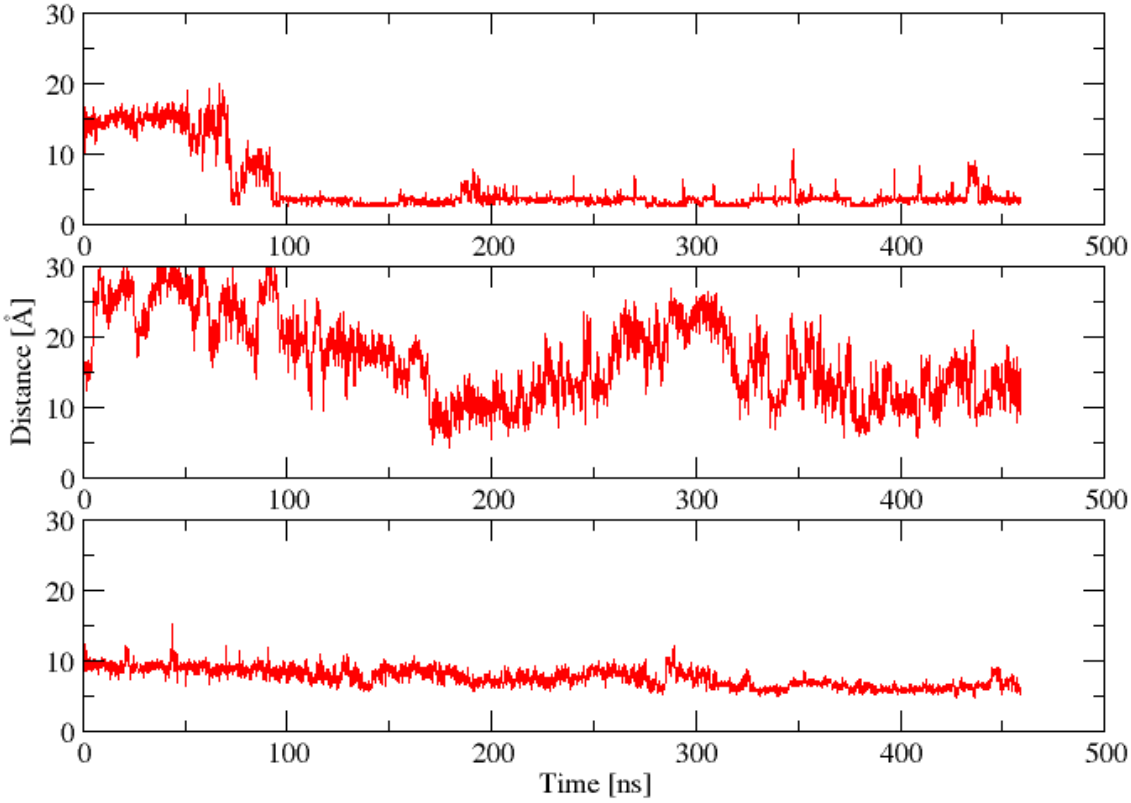


Figure S5: Newly formed salt bridges between TF and FVIIa in the simulation of the tenase complex.

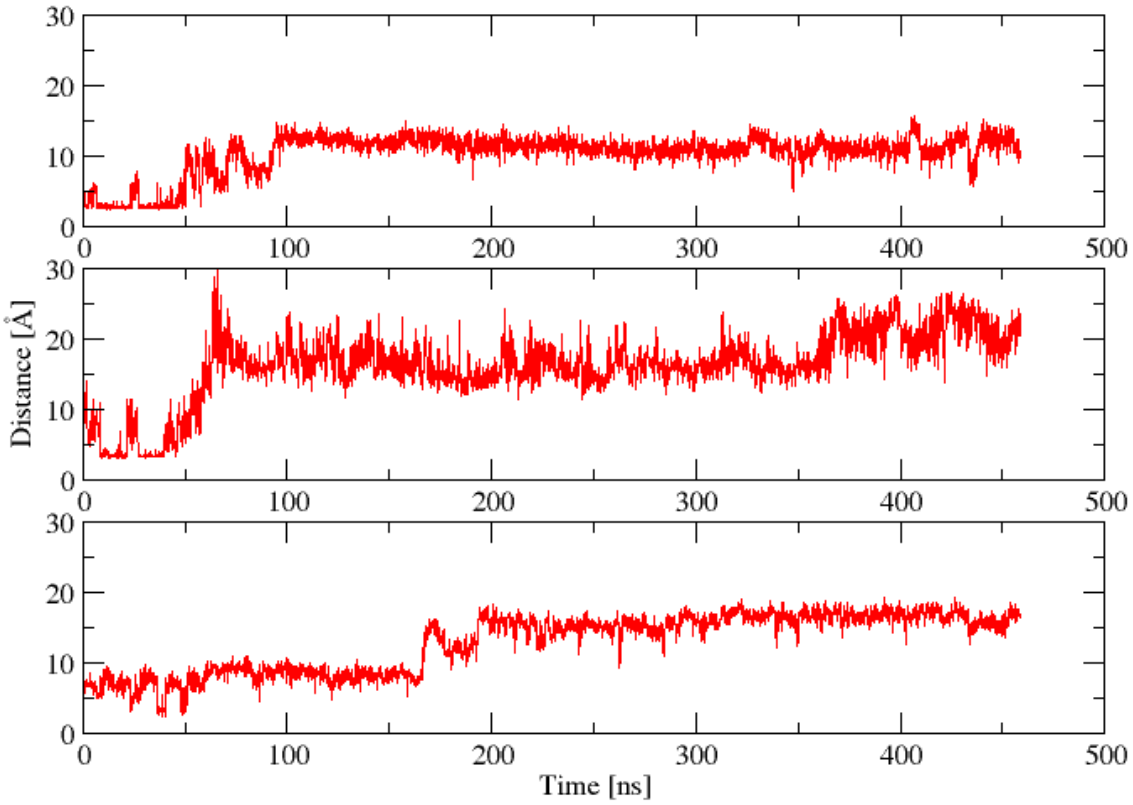


Figure S6: Broken salt bridges between TF and FVIIa in the simulation of the tenase complex.

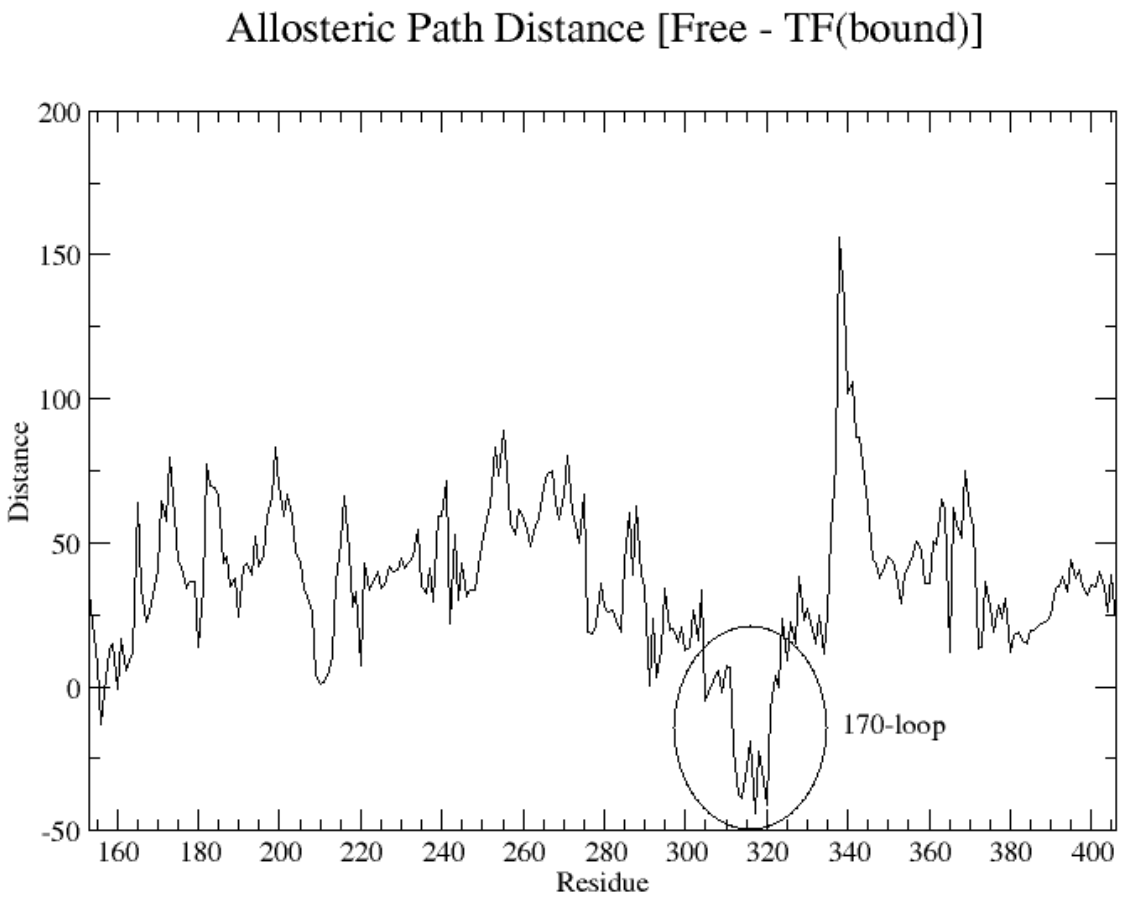


Figure S7: Trajectory-average distance of (shortest) allosteric paths from M306 to each other residue in the FVIIa protease domain. Plottet value reflects the difference between the free FVIIa and TF:FVIIa.

Path	*TF*	*Relative Frequency*	*Length [mean ± std.dev.]*
193-242-378-376-329-306	-	1	504.21 ± 56.6228
193-363-377-329-306	-	0.782609	513.824 ± 59.693
193-242-378-327-329-306	-	0.695652	509.219 ±

			57.1741
193-242-378-327-323-306	-	0.413043	504.439 ± 54.3949
193-191-361-377-329-306	-	0.369565	483.353 ± 53.3934
193-363-376-329-306	-	0.333333	475.609 ± 68.6191
193-362-377-329-306	-	0.318841	526.364 ± 60.9985
193-363-378-376-329-306	-	0.304348	518.667 ± 53.6837
193-242-327-329-306	-	0.289855	499.675 ± 61.85
193-242-326-323-306	-	0.275362	499.711 ± 66.8139
193-242-378-376-329-306	+	1	466.537 ± 55.952
193-363-377-329-306	+	0.744681	445.6 ± 57.211
193-363-376-329-306	+	0.56383	436.717 ± 49.1933
193-242-327-329-306	+	0.510638	454.583 ± 62.1238
193-242-378-327-329-306	+	0.5	468.894 ± 54.5591
193-242-378-328-306	+	0.297872	446.517 ± 59.1594
193-242-240-327-329-306	+	0.265957	500.16 ± 59.7339
193-364-376-329-306	+	0.255319	460.75 ± 65.1967

193-242-240-326-328-306	+	0.212766	481.7 ± 58.6985
193-242-327-323-306	+	0.191489	497.778 ± 54.7332

Table S1: Observed frequencies and calculated average lengths of the shortest allosteric paths between H193 and M306 elucidated by the sliding window analysis of the molecular dynamics trajectories with (+) and without (-) TF present. Top 10 most frequent paths are listed.

Path	*TF*	*Relative Frequency*	*Length [mean ± std.dev.]*
306-329-377-362-344	-	1	453.32 ± 62.3571
306-329-377-363-344	-	0.299595	471.351 ± 60.3862
306-329-376-378-362-344	-	0.242915	471.183 ± 53.1992
306-310-329-377-362-344	-	0.226721	482.464 ± 66.2309
306-323-327-378-362-344	-	0.153846	483.105 ± 54.9199
306-329-327-378-362-344	-	0.1417	475.943 ± 52.4437
306-323-327-377-362-344	-	0.117409	489.759 ± 44.9188
306-329-377-361-346-344	-	0.11336	481.179 ± 64.7512
306-329-376-363-344	-	0.11336	450.036 ±

			66.7969
306-329-376-362-344	-	0.105263	436.654 ± 69.2503
306-329-377-362-344	+	1	391.764 ± 48.3909
306-329-377-363-344	+	0.68125	392.339 ± 62.14
306-329-376-363-344	+	0.51875	384.181 ± 62.7698
306-329-376-364-344	+	0.3125	395.56 ± 43.3839
306-328-377-362-344	+	0.2375	383.421 ± 51.39
306-328-378-362-344	+	0.2	382.375 ± 43.0521
306-310-329-377-362-344	+	0.2	429.625 ± 47.8433
306-329-376-378-362-344	+	0.13125	432.143 ± 48.9424
306-329-376-362-344	+	0.125	388 ± 25.3791
306-329-327-363-344	+	0.10625	398.647 ± 78.8035

Table S2: Observed frequencies and calculated average lengths of the shortest allosteric paths between M306 and S344 elucidated by the sliding window analysis of the molecular dynamics trajectories with (+) and without (-) TF present. Top 10 most frequent paths are listed.

Path	*TF*	*Relative Frequency*	*Length [mean ± std.dev.]*

153-300-302-304-306	-	1	422.871 ± 52.7176
153-338-375-329-306	-	0.451613	434.822 ± 57.5339
153-339-375-329-306	-	0.435484	440.709 ± 56.8498
153-338-331-304-306	-	0.41129	422.112 ± 61.5784
153-281-302-304-306	-	0.41129	454.474 ± 49.1653
153-338-374-329-306	-	0.354839	386.692 ± 68.0321
153-340-366-376-329-306	-	0.274194	437.96 ± 43.481
153-338-374-310-306	-	0.25	471.8 ± 54.3831
153-339-377-329-306	-	0.233871	440.889 ± 74.6934
153-339-366-376-329-306	-	0.233871	412.077 ± 33.5432
153-300-302-304-306	+	1	395.121 ± 43.4256
153-281-302-304-306	+	0.99005	386.053 ± 47.8504
153-302-304-306	+	0.58209	384.759 ± 45.2953
153-377-329-306	+	0.487562	387.686 ± 52.6712
153-281-279-304-306	+	0.383085	395.484 ± 52.0853
153-377-327-323-306	+	0.129353	399.409 ±

			49.6869
153-335-375-329-306	+	0.124378	404.795 ± 43.8238
153-283-281-302-304-306	+	0.124378	404.258 ± 40.0266
153-377-329-310-306	+	0.0895522	405.414 ± 43.6076
153-281-279-302-304-306	+	0.0646766	406.062 ± 46.1986

Table S3: Observed frequencies and calculated average lengths of the shortest allosteric paths between I153 and M306 elucidated by the sliding window analysis of the molecular dynamics trajectories with (+) and without (-) TF present. Top 10 most frequent paths are listed.

Path	*TF*	*Relative Frequency*	*Length [mean ± std.dev.]*
306-323-321	-	1	247.481 ± 61.804
306-311-321	-	0.677249	242.734 ± 57.3958
306-309-311-321	-	0.518519	250.602 ± 54.1262
306-308-311-321	-	0.486772	247.239 ± 59.4945
306-309-313-315-317-319-321	-	0.470899	291.393 ± 52.8648
306-309-312-314-317-319-321	-	0.460317	295.333 ± 48.1785
306-311-317-319-321	-	0.439153	299.41 ± 48.8474

306-310-314-317-319-321	-	0.439153	299.771 ± 46.6583
306-310-317-319-321	-	0.428571	306.481 ± 49.4229
306-309-314-317-319-321	-	0.402116	296.592 ± 48.72
306-310-323-321	+	1	295.75 ± 44.6161
306-323-321	+	0.953125	246.213 ± 41.0298
306-311-321	+	0.9375	242.583 ± 47.7854
306-329-323-321	+	0.90625	288.914 ± 43.4493
306-309-311-321	+	0.765625	250.816 ± 52.6683
306-308-311-321	+	0.765625	244.265 ± 50.6709
306-329-327-321	+	0.640625	278.415 ± 37.3735
306-329-376-321	+	0.46875	275.6 ± 43.0439
306-329-327-323-321	+	0.40625	295.038 ± 49.8355
306-310-307-323-321	+	0.359375	313.957 ± 46.285

Table S4: Observed frequencies and calculated average lengths of the shortest allosteric paths between M306 and P321 elucidated by the sliding window analysis of the molecular dynamics trajectories with (+) and without (-) TF present. Top 10 most frequent paths are listed.

Path	*TF*	*Relative Frequency*	*Length [mean ± std.dev.]*
306-329-377-362-364	-	1	435.315 ± 49.4386
306-329-376-366-364	-	0.75	418.556 ± 53.6645
306-310-317-366-364	-	0.694444	383.8 ± 64.9896
306-329-376-364	-	0.675926	326.644 ± 58.5092
306-311-317-366-364	-	0.62963	392.574 ± 62.9835
306-309-313-315-317-366-364	-	0.37963	417.829 ± 59.0349
306-310-314-317-366-364	-	0.37037	391.225 ± 63.8985
306-329-376-378-364	-	0.361111	440.308 ± 34.7195
306-309-312-314-317-366-364	-	0.351852	391 ± 66.8069
306-308-311-317-366-364	-	0.342593	394.081 ± 69.0532
306-329-376-364	+	1	318.489 ± 55.0759
306-329-376-366-364	+	0.369318	363.6 ± 46.7809
306-329-375-366-364	+	0.221591	372.231 ± 51.3172
306-310-329-376-364	+	0.221591	362.308 ± 54.5138

306-329-327-364	+	0.198864	377.286 ± 52.2721
306-329-377-362-364	+	0.142045	399.16 ± 50.2201
306-328-376-364	+	0.119318	333.238 ± 48.3075
306-329-377-364	+	0.0965909	345.647 ± 65.2046
306-310-376-364	+	0.0909091	348.688 ± 49.2211
306-310-374-366-364	+	0.0738636	371.154 ± 33.9158

Table S5: Observed frequencies and calculated average lengths of the shortest allosteric paths between M306 and W364 elucidated by the sliding window analysis of the molecular dynamics trajectories with (+) and without (-) TF present. Top 10 most frequent paths are listed.

3.2.3 Notes: Trypsin 170-loop-swapped FVIIa variants

Introduction

In these notes, an activity-governing loop segment from the structural ancestor, trypsin, is integrated into FVIIa to engineer a protease that is independent of its biological cofactor, tissue factor (TF). Structural and functional characterization of the designed FVIIa variants is employed to yield novel insights into the underlying allosteric mechanisms inherent to the scaffold of this important protease.

Several of the trypsin-like proteases are poor enzymes after proteolytic activation and require cofactors to be significantly active. This fundamental theme enables the high degree of regulation as required in for instance emergency responses; typical examples include common coagulation factors (VIIa, IXa) and complement factors (B, C2, and D).(Gohara & Di Cera, 2011) Also the ubiquitous thrombin, albeit somewhat active after final processing of prothrombin, is further stimulated by cofactors. These observations suggest that (activated) trypsin-like proteases coexist in molecular states which can be either active or inactive and, furthermore, that the trypsin-fold is capable of supporting enzymatic allostery. Capturing these mechanisms, however, is a major challenge because native states (wild-type, free and uninhibited) are difficult to crystallize.

The 170-loop of FVIIa is peculiar in that is five residues longer in FVIIa than in the ancestral trypsin and also related proteases from the coagulation system (thrombin, FIXa, FXa, PC). A multiple sequence alignment show that this characteristic encompasses the largest inserted region (disregarding the C-ternimally positioned residues) of the SP domain of FVIIa. This observation is interesting since the primary point of origin for TF-induced allostery in FVIIa is likely M306(Persson, Nielsen, & Olsen, 2001)(Rand et al., 2008), positioned at the N-terminal part of the TF-binding helix and directly leading into the 170-loop. Accumulating evidence based on mutagenesis studies have elaborated two principal allosteric pathways in FVIIa originating from the TF-binding helix.(Persson & Olsen, 2011) The associated mechanistic pictures are rationalized on the basis of X-ray crystallographic structures. One signal of allostery (path I) travels up through the 170-loop of FVIIa, which is stabilized in the binary complex with TF, and the P321 ‘clamps’ down on the W364, assisting the maturation of substrate-binding subsites. In an orthogonal

direction, the conformation of activation loop 3 is altered in a manner which facilitates appropriate N-terminal insertion into the activation pocket, forming the hallmark salt bridge interaction with D242{c194} (path II). Interestingly, grafting the 170-loop from trypsin seems to short-circuit both pathway and modify the behavior of FVIIa as to become constitutively active in the absence of TF.(Soejima et al., 2001b)

Conversely, certain functionalities from arguably more complicated trypsin-like proteases have been successfully integrated back into trypsin, e.g. Na^+ allostery based on insertion of thrombin fragments(Page & Di Cera, 2010) and the modulation of inhibitor specificity taken from FXa(Rauh, Klebe, & Stubbs, 2004).

In the present notes, we demonstrate that it is the flexible nature of FVIIa that requires it to be tamed by its cofactor. However, the shortening of the 170-loop alone does not accomplish this; a key residue interaction is formed by Y314{c170B} cause the segment flanking the active site (residues 364{c215}-368{c219}) to be stabilized. These findings are, to the best of our knowledge, the first indication that tryptophan-gated allosteric mechanisms are present in FVIIa.

Experimental

Preliminary experimental data (the following two tables) for reference and comparison of the presented FVIIa variants was graciously provided by Anders B. Sørensen (to be published).

Table 4 *sTF binding to the FVIIa variants using surface plasmon resonance (SPR).*

	Func. K_d [nM]	Steady-State SPR K_d [nM]	SPR k_{on} [1/Ms]	SPR k_{off} [1/s]
FVIIa-WT	1.2 (0.3)	6.1 (1.4)	448971	0.0027 (0.0004)
FVIIa-311	101.6 (4.1)	140.7 (8.3)	213057	0.030 (0.0023)
FVIIa-313	34.2 (2.0)	38.0 (1.6)	293848	0.011 (0.0001)
FVIIa-860	231.1 (22.8)	248.7 (18.2)	180048	0.045 (0.0024)

Table 5 *Kinetic parameters for the hydrolysis of S2288, pABA inhibition values and carbamylation halftimes as means with calculated SEM, N=2.*

	S2288 K_m [mM]	S2288 V_{max} [mAu/min]	xWT[a]	pABA Ki [uM]	KNCO $T_{½}$ [min]
-sTF					
FVIIa-WT	8.9 (0.4)	135.5 (3.2)	1.0	1485 (87.7)	43.9 (1.0)
FVIIa-311	4.0 (0.1)	535.3 (8.5)	8.8	97.6 (10.1)	40.8 (0.5)
FVIIa-313	5.3 (0.2)	341.2 (5.1)	4.2	190.4 (13.1)	43.8 (0.6)
FVIIa-860	15.4 (1.3)	120.2 (6.4)	0.5	490.3 (46.3)	32.4 (0.7)
+sTF					
FVIIa-WT	1.2 (0.03)	610.5 (3.8)	32.7	49.3 (1.8)	536.2 (39.6)
FVIIa-311	1.2 (0.02)	735.1 (4.0)	40.0	24.7 (1.2)	56.7 (0.9)
FVIIa-313	3.0 (0.06)	653.3 (5.5)	14.6	41.7 (1.3)	159.0 (4.8)
FVIIa-860	5.6 (0.1)	381.5 (4.4)	4.5	69.5 (3.7)	37.1 (0.7)

[a] FVIIa-WT V_{max}/K_m arbitrarily set to one and all other values calculated as relative efficiency

Materials and Methods

Model construction of FVIIa variants.

The FVIIa 170-loop variants were constructed using the X-ray crystallographic structure of the 311-FVIIa variant as a template (to be published, L. A. Svensson). For the 313- and 860-variants, the tyrosine 314{c170B} was mutated to serine and phenylalanine respectively while preserving the sidechain rotamer from the template. WT-FVIIa was based on a representative structure of benzamidine-inhibited FVIIa ((Sichler et al., 2002), PDB entry code 1kli). The complete complex of soluble Tissue Factor (residues 1-213) and FVIIa was constructed starting from PDB entry code 1dan as described in ((Y Z Ohkubo et al., 2010), and graciously provided by the authors) where the absent loops were taken from other PDB entries (see also ms.2). For all structures, the co-crystallized inhibitor was removed.

Simulation details.

Conventional molecular dynamics (MD) simulations of the WT-FVIIa:sTF complex, WT-FVIIa, and the three FVIIa variants with the trypsin 170-loops were performed using the NAMD 2.7 software package(Phillips et al., 2005) with the CHARMM27 force field(MacKerell et al., 1998) and the TIP3P water model(Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) similarly to as described in ms.1 and ms.2 previously.

Multiple Sequence Alignment.
Retrieval and alignment of the 170-loop sequences for trypsin and FVIIa was done using UniProt(The Uniprot Consortium, 2013) and the Clustal Omega algorithm(Sievers et al., 2011).

Structural modeling of 170-loop conformations of the 311-, 313-, and 860-FVIIa variants.
Conformations of the 170-loop in the scaffold of the X-ray crystallographic structure: Sampling of the loop conformational space was done using the Rosetta modeling suite(Leaver-Fay et al., 2011). Scores (pseudo-energies) of the generated decoy models are evaluated under the standard score12 scorefunction unless otherwise stated. A scaffold for examining the loop conformations of the 311-, 313-, and 860-FVIIa variants was created by removing the 7 variant-specific residues EASXPGK (Fig. 13) in the X-ray crystallographic structure resolved for the 311-variant (to be published, L. A. Svensson). This scaffold was processed by a high-resolution protein structure refinement protocol with backbone perturbations and sidechain repacking (Fast Relax protocol, unpublished). 1,200 decoys were generated and the best scoring decoy was chosen as a representative model for further modeling. On this scaffold, the 170-loop was reconstructed and refined for each of the three loop sequences for 311-FVIIa, 313-FVIIa, and 860-FVIIa, respectively, by the kinematic closure (KIC) protocol in Rosetta(Mandell, Coutsias, & Kortemme, 2009). >20,000 loops were modeled for each 170-loop variants and the best scoring (top 10 %) formed the representative ensemble of accessible loop conformations, and were grouped by structural clustering.

Results and Discussion

The N-terminus salt bridge between the amino group of I153{c16} and D343{c194} is formed throughout the entire duration of all simulated FVIIa variants; The trajectory

averaged distance between the N atom of the amino group and the C_γ atom of the carboxylic acid was 3.21-3.29 Å, indicating a stable salt bridge. Since carbamylation is the default assay for distinction between the zymogen-like FVIIa and the functionally active FVIIa population, we here adopt the notion, that the correctly inserted N-terminus is *the* structurally defining marker for the enzyme, and thus, any activity regulating conformational changes distinguish between the active enzyme (E) and the inactive enzyme (E*). Sequence alignment of the stidued FVIIa variants with WT-FVIIa and trypsin can be seen in the figure below (Fig. 13).

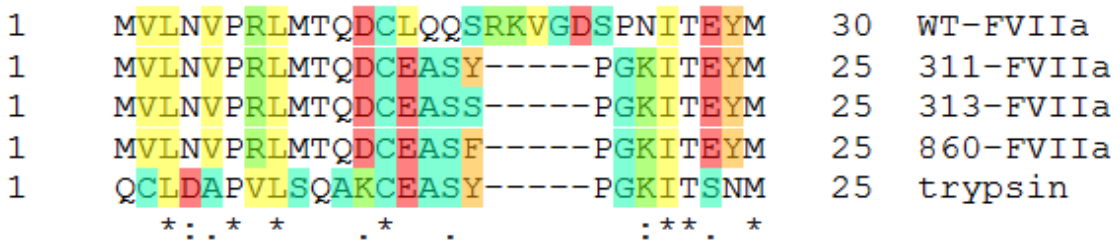


Figure 13 *Multiple sequence alignment of the 170-loops of the loop-swapped FVIIa variants.*

Dynamics Simulations of the FVIIa variants.

Tryptophan-364{c215} is located in between activation loops 2 and 3 of the SP domain of FVIIa (Fig. 4B). In the X-ray crystallographic structures of inhibited FVIIa, well represented by PDB entry code 1kli(Sichler et al., 2002), W364{c215} hydrogen bonds with V376{c227} as part of a beta-sheet of three juxtaposed beta-strands forming one side of the twin beta-barrel domain topology. Furthermore, the amino acid residues surrounding W364{c215} constitutes one rim of the S1 substrate specificity pocket. The β-barrel region of the SP domain of FVIIa which harbors activation loop 1, activation loop 2, activation loop 3, and the 170-loop is more disordered than the opposing half of the domain. In our molecular dynamics simulations all three activation loops undergoes significant structural changes, and of particular interest, so does the region following the E2 strand sheet which harbors W364{c215}. This is expected of surface exposed loops, but the difference in flexibility between WT-FVIIa and the sTF:WT-FVIIa binary complex is astounding. The re-arrangements of loops is most pronounced in WT-FVIIa where W364{c215} not only releases from the aryl binding pocket S3/S4 (Fig. 14B), but the S1 pocket collapses entirely, as indicated by the short distance between the Tryptophan and the S344{c195}

(Fig. 14C). The molecular dynamics simulations clearly show that TF effectively tame the dynamics of WT-FVIIa, keeping it in the productive "open" (E) conformation which is structurally and dynamically similar to the active-site inhibited FVIIa.

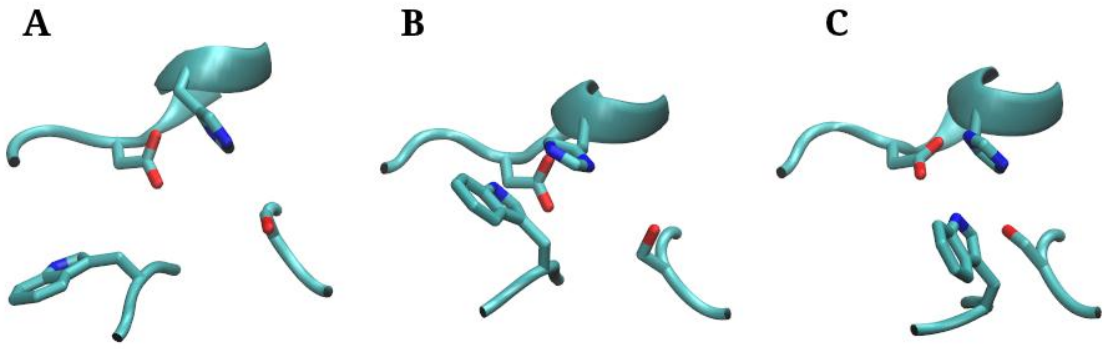


Figure 14 *Representative structural conformations of Tryptophan 215{c} relative to the catalytic triad and scatter plots of the distances between W215{c} and the active H57{c}/S195{c} in the molecular dynamic simulations of the FVIIa variants.*

*Snap-shots were taken from the simulation of WT-FVIIa. (**A**) crystal structure and "open" (E) conformation. (**B**) E* conformation of W215{c}. (**C**) E* conformation of W215{c} and collapsed S1 pocket.*

Most remarkably, this conformational transition to an unproductive (E*) population is likewise absent in the trypsin 170-loop variant of FVIIa without Tissue Factor, but is re-introduced in the 313- and 860-FVIIa variants by modifying a single key stabilizing polar interaction of the hydroxyl group of the trypsin 170-loop Y314{c170B} (Fig. 15). The fact that the Y314F mutation of 860-FVIIa in the trypsin 170-loop scaffold has such a dramatic reversing effect reveals that stacking alone of the bulky hydrophobic sidechains is not sufficient to significantly bias W364{c215} toward the "open" (E) conformation.

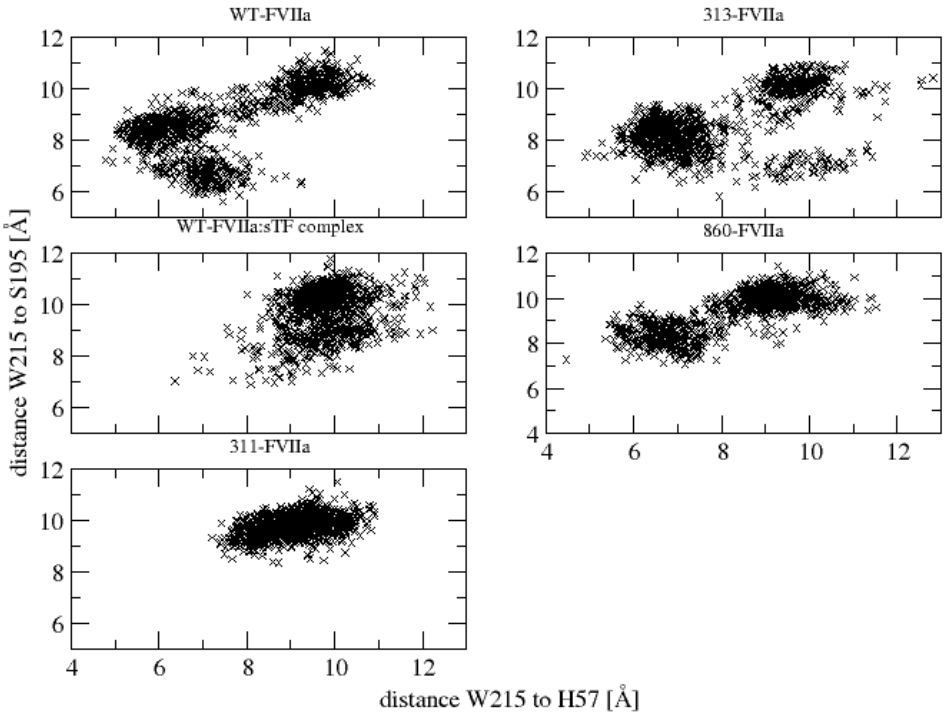


Figure 15 *Scatter-plots of W215{c} and the active H57{c}/S195{c} distances calculated every 1 ns during the 100 ns-simulations.*

170-Loop Conformational Sampling.

Visual inspection and comparison of crystal structures for WT-FVIIa and 311-FVIIa suggests that the P315{c170C} in the trypsin 170-loop variants may restrain the motion of the W364{c215} by lidding off the indole group and keeping it in the "open" conformation. The 170-loop conformational search for the FVIIa variants 311, 313, and 860 can be seen in Fig. 16, where the pseudo-energy of the best scoring loop decoys are plotted against their geometric distance (RMSD) to the native 311-FVIIa 170-loop. For each of these three variants, 5-6 major populations of loop conformations are observed. Surprisingly, though, pseudo-energy landscapes looks indistinguishable between 311- and 860-FVIIa variants, whereas 313-FVIIa is notably different.

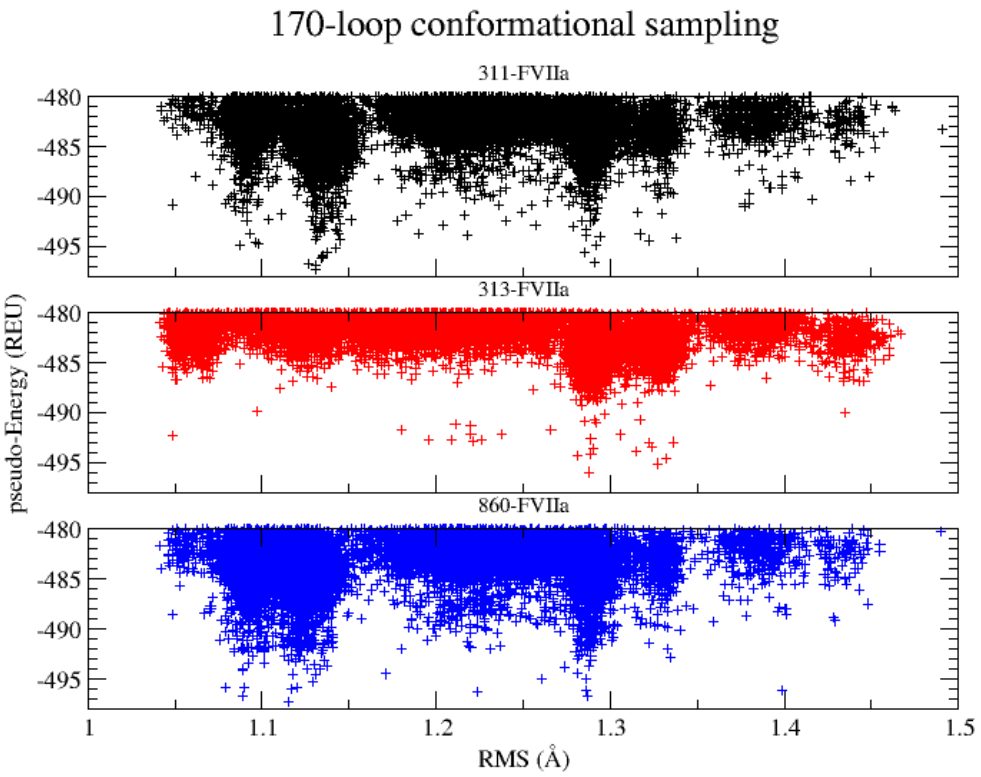


Figure 16 *pseudo-Energy vs geometric distance of the generated 170-loop decoys for the 311-, 313-, and 860-FVIIa variants on the 311-FVIIa scaffold.*

Concluding remarks

The position of the W364{c215} is crucial for the structural integrity of the aryl-binding site of the substrate binding cleft in FVIIa. In the vast majority of published x-ray crystallographic structures of inhibited FVIIa, W364{c215} is aligned to a single conformation, well represented by PDB entry code 1kli(Sichler et al., 2002). Tryptophans which can exist in various structural conformations are known to be important in allosteric regulation pathways in other trypsin-like serine proteases, e.g. Thrombin(Gandhi, 2010). The conformational transition seen in our simulations has been reported previously for the homologous factor Xa in a computational study (Wang, Hao, & Li, 2011). There is strong evidence that the Tryptophan 215 {364} conformational change observed in our molecular dynamics for the preferentially “open” FVIIa variants is indeed physically attainable ((Rai et al., 2006),PDB entry code 2f9b). Though, to the best of our knowledge, the present

study is the first indication of the critical role of W364{c215} in FVIIa allostery both upon TF-binding and general activity modulation in the trypsin 170-loop-swapped variants.

3.3 TF:FVIIa complex formation

3.3.1 Prelude

Manuscript 3 describes a functional and structural characterization of FVIIa variants with truncations of increasing length (one to four residues) in the inter-domain linker between the two EGF-like domains. Said linker has a defining role in the ability of FVIIa to interact with TF because it works as a conformational hinge that can accommodate rotated light chain conformations (Fig. 5). It is intriguing to note that this functionality is apparently preserved even when one or two residues are removed (starting at position residue 86 and deleting toward the N-terminus), still allowing a productive interaction with TF to be formed. However, truncation by three or four residues completely abolished TF-mediadated stabilization and, consequently, TF-induced activity enhancement.

Simulations of the free FVIIa linker variants substantiate that no important structural elements are destroyed by the linker truncations up till 3 residues toward the N-terminus from position 86. Furthermore, enforcing the FVIIa linker variants to complex with TF shows that this interaction becomes energetically un-favourable when exactly three (or four) residues are removed, but not when only one or two residues are removed. These results are in excellent agreement with experimental data.

3.3.2 Manuscript 3: Persson E, *et al.*, submitted

The Length of the Linker between the Epidermal Growth Factor-like Domains in Factor VIIa Is Critical for a Productive Interaction with Tissue Factor

Egon Persson*[,1], Jesper J. Madsen[2] and Ole H. Olsen[2]

[1]Haemophilia Biology and [2]Haemophilia Biochemistry, Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark.

*Correspondence: Egon Persson, Haemophilia Biology, Novo Nordisk A/S, Novo Nordisk Park, F5.2.11, DK-2760 Måløv, Denmark; tel.: 45-30754351; e-mail: egpe@novonordisk.com

Running title: FVIIa linker determines tissue factor complementarity

ABSTRACT: Formation of the factor VIIa (FVIIa)-tissue factor (TF) complex triggers the blood coagulation cascade. Using a structure-based rationale, we investigated how the length of the linker region between the two epidermal growth factor (EGF)-like domains in FVIIa influences TF binding and the allosteric activity enhancement, as well as the interplay between the γ-carboxyglutamic acid (Gla)-containing and protease domains. Removal of two residues from the native linker was compatible with normal cofactor binding and accompanying stimulation of the enzymatic activity, as was extension by two (Gly-Ser) residues. In sharp contrast, truncation by three or four residues abolished the TF-mediated stabilization of the active conformation of FVIIa and abrogated TF-induced activity enhancement. In addition, FVIIa variants with short linkers associated 80-fold slower with soluble TF (sTF) compared with wild-type FVIIa, resulting in a corresponding increase in the equilibrium dissociation constant. Molecular modeling suggested that the shortest FVIIa variants would have to be forced into a tense and energetically unfavorable conformation in order to be able to interact productively with TF, explaining our experimental observations. We also found a correlation between linker length and the residual intrinsic enzymatic activity of Ca^{2+}-free FVIIa; stepwise truncation resulting in gradually higher activity with des(83-86)-FVIIa reaching the level of Gla-domainless FVIIa. The linker appears to determine the average distance between the negatively charged Gla domain and a structural element in the protease domain, presumably of opposite charge, and proximity has a negative impact on apo-FVIIa activity.

Keywords: Factor VIIa; tissue factor; allostery; structural complementarity; linker region

Statement for the broader audience: A linker region connects the N- and C-terminal domain pairs in factor VIIa (FVIIa) and other coagulation factors. FVIIa tolerates a deletion of only two amino acid residues before the complementarity with tissue factor (TF) is lost and cofactor function is abolished. This illustrates the pivotal spacer function of a linker region, an optimal length being crucial for a productive, high-affinity interaction. Linker regions dictate the distance between components of composite, multi-domain binding interfaces and presumably play important roles in numerous protein-protein interactions.

Abbreviations: FVII(a), (activated) coagulation factor VII; TF, tissue factor; sTF, soluble TF (residue 1-219); EGF, epidermal growth factor; Gla, γ-carboxyglutamic acid.

Coagulation factor VIIa (FVIIa)[3] is a serine protease of pivotal importance to the blood coagulation cascade, forming the initiating complex with its cognate cofactor, tissue factor (TF) (1). The precursor FVII is synthesized as a single-chain, inactive zymogen, and FVIIa results from activation by endoproteolytic cleavage after Arg-152. Trace amounts FVIIa patrol the vasculature and respond instantly to the exposure of TF as a sign of vascular damage. Free FVIIa is an enzyme in a quiescent state lacking significant biological activity, but upon association with TF it becomes an efficient catalyst capable of rapidly initiating the blood coagulation cascade by converting more FVII to FVIIa and by activating factors IX and X (1). The TF-interactive interface on FVIIa (2) can be divided into two parts, the first involving the light chain and providing the majority of the binding energy (3) and the second primarily being composed of contact points on the protease domain, including Met-306 which is the origin of the allosteric effect on FVIIa (4). FVIIa is homologous to other vitamin K-dependent coagulation factors and contains the same four domains as, for instance, factors IX and X, namely a γ-carboxyglutamic acid (Gla)-containing domain, two epidermal growth factor (EGF)-like domains and a serine protease domain (5). The first three domains make up the light chain, whereas the protease domain constitutes the heavy chain. The last Cys residue of the first EGF-like domain and the first Cys residue of the second EGF-like domain are separated by a 9-residue linker sequence. The linker comprises a hinge region (one residue in particular) which permits different relative orientations of the domain pairs on each side (Gla plus first EGF-like domains and second EGF-like plus protease domains, respectively) (6). In the complex with TF, FVIIa has an extended conformation with a stretched linker region dictated by the interactions with the protein cofactor (2).

Even though the amino acid sequence of the linker in FVII(a) is not conserved between species (7, 8), its length and flank regions, including the most pronounced hinge residue Gln-88, are. This suggests that interdomain distance and flexibility are of functional importance. Factors IX(a) and X(a) on the other hand have a shorter linker consisting of five and seven amino acid residues, respectively. Only the first glutamic acid residue is conserved between the three human factors (9,

10). Rather than directly participating in protein-protein interactions, the linker regions may serve to define the distance between distant parts of the proteins and allow them to interact optimally with for instance protein cofactors and substrates.

In this study, we execute a structure-based truncation strategy to demonstrate that the length of the linker between the EGF-like domains in FVIIa is crucial for the optimal physical interaction with TF. This is in turn necessary for stabilization of the active conformation of FVIIa and a productive outcome in terms of enhanced enzymatic activity. Too short a linker cannot, at least not in any low-energy conformation with allowed backbone dihedral angles, support an appropriate distance between the two EGF-like domains nor between distant patches of the TF-interactive interface on FVIIa. We also propose that the linker length governs the spatial distance between the Gla domain and an element in the protease domain and thereby influences the residual enzymatic activity of the apo form of free FVIIa.

RESULTS

Molecular design rationale behind linker truncations

Alignment of FVIIa sequences from a wide range of mammals reveals that linker residues Glu-82, Gln-88, Leu-89, and Ile-90 are invariably conserved, whereas positions 83-87 are promiscuous with respect to amino acid occupancy. This strongly suggests that deletions within residues 83-87 will affect the linker length without disrupting important secondary structural elements. Compared with the X-ray crystal structure of FVIIa bound to TF (2), structures of free FVIIa with visible EGF-like domains (6, 26; PDB entry codes 1qfk and 1dva) confirm the presence of the short β-strand at residues Thr-83 to His-84, but reveal no consensus on the position of the helical fragment observed. This suggests that the flexible nature of the linker allows the formation of half or full turns but only transiently, consistent with previous computational studies of free and TF-bound FVIIa (27). We conclude that only the length of the linker is likely to be affected in the des86- and

des(85-86)-FVIIa variants, whereas in des(84-86)-FVIIa, Asp-87 will move to the position occupied by the native residue 84. This creates a sequence (Glu-82, Thr-83, Asp-84) present in bovine and canine FVIIa making retention of the β-sheet in free des(84-86)-FVIIa plausible. However, in the structural minimizations of the des(84-86)-FVIIa-TF complex, no β-sheet is observed. This is not due to the relative β-sheet propensity of aspartic acid, but because the linker length no longer allows it. The β-sheet might be entirely missing in free des(83-86)-FVIIa. In order to evaluate the physical extension of various linker lengths, molecular dynamics simulations were performed on native FVIIa and the four truncated variants. Representative snapshots from the simulations are shown in Fig. 1. The distance between the last disulfide in the first EGF-like domain and the first disulfide of the second EGF-like domain decreases with the number of residues truncated while, indeed, the helical element observed in the X-ray crystal structure of FVIIa-TF (2) is missing in all the simulated free FVIIa variants and, instead, residues 85-91 are engaged in varying degrees of intermittent turn/coil transitioning (Fig. 1). In addition, the short β-sheet in which strand residues Thr-83 and His-84 participate is preserved in the simulation of FVIIa. The strand property is retained in simulations of the truncated des86-, des(85-86)-, and des(84-86)-FVIIa variants, but is absent in des(83-86)-FVIIa. These results suggest that the linker region in all the free FVIIa variants is devoid of helical elements and that a short strand is present when up to three residues are removed. Hence truncation of the linker, at least by up to three residues, affects its length but not secondary structural content.

Effects of sTF on activity and conformational stability of FVIIa linker variants

One purpose of this study was to investigate whether the length of the linker between the two EGF-like domains in FVIIa is critical for the interaction with, and cofactor-induced effects of, TF. In the initial characterization, all the FVIIa variants, both those with an extended and those with a truncated linker, were found to possess virtually the same intrinsic amidolytic and proteolytic activity as wild-type FVIIa (data not shown). The latter was measured using FX as the substrate

both in the presence and absence of phospholipids. In the next step, the amidolytic activities were measured in sTF titration experiments. The FVIIa variants with the smallest truncations, lacking one (des86-FVIIa) or two residues (des(85-86)-FVIIa) in the linker, responded normally to sTF with a fold activity enhancement at saturating concentrations of sTF and an EC_{50} value indistinguishable from those of wild-type FVIIa (Fig. 2A). In sharp contrast, those with longer truncations, i.e. des(84-86)- and des(83-86)-FVIIa, were virtually non-responsive to sTF (even at 5 μM cofactor). An increased incubation time with sTF did not help to activate the two shortest FVIIa variants (data not shown).The variants containing extensions of the linker, 86G87- and 86GS87-FVIIa, behaved like FVIIa (Fig. 2B).

Carbamylation, a chemical modification of the N-terminal amino group of the protease domain and a measure of its accessibility to solvent (28), results in loss of enzymatic activity. It was used to investigate whether the ability (or inability) of sTF to enhance the activity of a particular FVIIa variant was reflected in an increased (or unaltered) degree of burial of the N-terminus in the presence of the cofactor. In the absence of sTF, all the studied FVIIa variants, whether truncated or extended in the linker, Gla-domainless or full-length, were modified at indistinguishable rates and lost 13-14% activity per 10 minutes. When bound to sTF, FVIIa, 86G87-FVIIa, 86GS87-FVIIa, des86-FVIIa and des(85-86)-FVIIa were all modified at a significantly slower rate (2-3%/10 min). In sharp contrast, the rates of carbamylation of des(84-86)- and des(83-86)-FVIIa were not affected by complex formation with sTF. Thus there appears to be an absolute correlation amongst the FVIIa linker variants between sTF-induced activity stimulation and burial of the N-terminus into the activation pocket.

Binding of FVIIa linker variants to sTF

The physical binding of FVIIa variants to sTF was examined with the main purpose to investigate whether the failure of sTF to stimulate certain FVIIa variants (those with the largest truncations of the linker region) was due to an abolished interaction. The kinetics of sTF binding were very

similar for FVIIa, 86GS87-FVIIa, des86-FVIIa and des(85-86)-FVIIa, albeit with slightly slower association rate and faster dissociation rate for des(85-86)-FVIIa (Fig. 3 and Table 1). These FVIIa variants also exhibited indistinguishable degrees of sTF-induced activity enhancement. In contrast, des(83-86)- and des(84-86)-FVIIa bound sTF with dramatically different kinetics due to the association rates being approximately 80 times slower than those of the longer variants. However, once bound, the complexes of des(83-86)- and des(84-86)-FVIIa with sTF were virtually as stable as those of the FVIIa variants with a longer linker region. This suggest that an interactive interface is eventually formed which provides the same binding energy (affinity) but without a concurrent stimulation of the catalytic machinery located in the protease domain.

Modeling of FVIIa linker variants in complex with TF

For a functional FVIIa-TF complex to be formed, FVIIa must attain a conformation which is complementary to TF and accommodate specific exosite interactions with TF. Our striking experimental finding that the entire change of the functional properties of FVIIa occurred upon removal of the third residue from the linker asked for a plausible explanation. In order to investigate how the length of the linker connecting the EGF-like domains in FVIIa affects the ability of the molecule to form a fully functional complex, the linker was gradually truncated by one to four residues and an attempt to structurally explain the slow, non-productive TF binding of des(83-86)- and des(84-86)-FVIIa was made by generating structural ensembles for all of the Ca^{2+}-loaded FVIIa variants in the TF-bound form. The nature of the linker in the X-ray crystallographic structure of FVIIa bound to TF (2) encompasses both a β-sheet, pairing the two-residue β-strand composed of Thr-83 and His-84 with Phe-76 and Glu-77, and a short 3_{10} helical fragment from Lys-85 to Asp-87. This suggests that partial unfolding of the helical element may make it possible for the FVIIa-TF complex to tolerate small deletions in the linker region with retained functionality, which is indeed what was experimentally observed.

The structural characterization of the phenomenon was done in two manners. Firstly, the strain in the X-ray crystal structure introduced by truncations in the linker was gauged. For this purpose, an ensemble of possible linker conformations were initially generated using Rosetta's loop modeling application on the otherwise fixed structure of the FVIIa-TF complex. This approach revealed that in the case of a four-residue deletion, no combination of allowed backbone dihedral angles and atomic bond lengths could yield a linker conformation connecting the two flanking Cys residues without a chain break (not shown). For this simple reason, we conclude that it is very unlikely that one can truncate the linker by four residues and retain the interactions with TF (at least those seen in the crystal structure) required for a functional FVIIa-TF complex. Therefore, a second approach to evaluate the consequences of linker truncation on the tertiary structure of the FVIIa-TF complex by forcing ideal geometry (and hence loop closure) was employed. If four residues were deleted and the bond lengths idealized, the FVIIa-TF complex adopts a warped conformation in which FVIIa clamps TF causing it to bend significantly and deviate from the known structure. Interestingly, the structural ensembles using the native linker or a linker after deletion of one or two residues scored virtually identically (Fig. 4). Inspection of representative structures selected by clustering methods showed that for a one- or two-residue deletion the loop straightens to some extent and the characteristic helical fragment of the linker indeed unfolds, but it does so without greatly penalizing the overall score of the complex or the linker region. On the other hand, structures for the three- and four-residue deletions were energetically punished, in part, by their unfavorable backbone dihedral angles which could be confirmed by Ramachandran plots for the linker residues (Fig. 4).

Influence of linker length and Gla domain on the activity of apo-FVIIa

In the presence of 5 mM Ca^{2+}, FVIIa and des(1-44)-FVIIa (Gla-domainless FVIIa) had similar amidolytic activities (K_m and k_{cat}) as measured with the substrate S-2288 (data not shown). NaCl in the range 0.1-0.3 M had a slightly stimulatory effect on the activity of both forms of FVIIa, an

effect which gradually decreased with increasing concentration (no net effect at 0.5 M) and turned inhibitory at higher concentrations (29).

In the absence of Ca^{2+}, full-length and des(1-44)-FVIIa exhibited much lower amidolytic activity, but des(1-44)-FVIIa retained significantly higher activity than FVIIa as published (30). The residual activities of the apo (Ca^{2+}-free) forms of FVIIa and des(1-44)-FVIIa were higher when the ionic strength was maintained with sodium chloride than with choline chloride (Fig. 5). Moreover, increasing the NaCl concentration (at least up to 0.6 M) gradually enhanced the activity of full-length and des(1-44)-FVIIa in the absence of Ca^{2+} (data not shown), whereas increasing the choline concentration had a marginal effect. This suggests that NaCl shields the charged Gla domain and further stimulates the amidolytic activity of FVIIa by saturation of a putative sodium site in the protease domain (31, 32).

With this background knowledge about the effects of various ions on and the properties of full-length and Gla-domainless FVIIa in the absence of Ca^{2+}, we studied the influence of the linker length. FVIIa retained about 7% amidolytic activity upon removal of Ca^{2+} from assay buffer containing 0.1 M choline chloride (Fig. 5). Similar residual activity was measured with 86GS87- and 86G87-FVIIa. As the linker length decreased, the FVIIa variants gradually retained more activity and des(83-86)-FVIIa possessed the same residual activity as des(1-44)-FVIIa (17-18%). The pattern was very similar when using NaCl as the salt, but the level of residual activity was generally twice as high and ranged from 13 to 35%, conceivably because of Na^{+}-specific stimulation (Fig. 5).

Carbamylation, even when using a relatively low concentration of potassium cyanate (50 mM) to achieve a decreased but still measurable rate of activity loss with FVIIa in the presence of Ca^{2+}, can apparently not distinguish between the degrees of N-terminus exposure in Ca^{2+}-loaded and -free FVIIa (data not shown).

DISCUSSION

Complex formation with TF is energetically driven by interactions involving the light chain of FVIIa, whereas establishment of the interactions involving the protease domain leads to allosteric stabilization of the active conformation of FVIIa (2-4). Together, this shapes a high-affinity, enzymatically active complex with FVIIa positioned for efficient substrate processing. In free FVIIa, the linker between the EGF-like domains in the light chain appears to be a hinge for interdomain flexibility (6, 27, 33). FVIIa bound to TF is motionally restricted and adopts an extended conformation which intuitively does not permit much reduction in linker length if the tight *and* productive interaction with TF, with participation of the light chain *and* protease domain, is to be maintained (2, 27, 33). Measurement of FVIIa amidolytic activity enhancement provides an unambiguous assessment of the stimulatory effect of TF looking exclusively at the catalytic machinery and the state of the S1 and proximal subsites without influence from exosites. We reduced the length of the linker by truncations at positions which, according to molecular dynamics simulations, minimally affected secondary structural elements. Importantly, the simulations of free FVIIa variants showed that the helical structure observed in FVIIa-TF is missing throughout. We were able to show that the TF-induced stimulation of FVIIa activity was abrogated when three amino acid residues are removed from the linker, whereas truncation of two residues, or an extension, were perfectly tolerated. The same, very dramatic pattern was observed for the physical interaction between FVIIa and sTF. Binding kinetics and affinity were indistinguishable from those of native FVIIa unless at least three residues were removed from the linker, in which case the association rate plummeted. These observations clearly demonstrated that a linker lacking three residues does not allow the physical separation of the light chain and protease domain parts of the TF-interactive interface to render allosteric activity enhancement and native affinity possible. The weakest part of the interface, that between the protease domain and TF, apparently capitulated and this resulted in loss of TF-induced activity increase. Finally, we found that a truncation of the linker

by four residues eliminated the negative impact of the Gla domain on the intrinsic activity of FVIIa in the absence of Ca^{2+}.

The slower sTF association rates of des(83-86)- and des(84-86)-FVIIa might be explained by their preferential existence in energetically favorable conformations encompassing allowed dihedral angles. However, because of the relatively short linkers in these variants, these conformations do not allow a productive interaction with sTF and establishment of the contacts seen in the crystal structure of the complex. The conformations which would fit to TF and form all the interactions known from the structure are, if at all possible, extremely unfavorable and scarce. This would represent a particular problem if FVIIa initially docks with the protease domain, in itself a low-affinity interaction with sTF. In this case, an imminent clash or a suboptimal distance between the two parts of the TF-interactive interface on FVIIa (protease domain and Gla/first EGF-like domains) would prevent the ensuing tethering of the complex by the contacts between the light chain of FVIIa and TF. This binding pathway has been inferred from transition state analysis and association rate studies of the FVIIa:sTF complex formation (34, 35). There is evidence in support of the notion that Gln-88 acts as hinge residue in the linker between the EGF-like domains in FVIIa (6). The first EGF-like domain has been resolved by X-ray crystallography in several structures of FVIIa in complex with TF with a similar domain configuration, well represented by the structure of Banner *et al.* (2). To our knowledge, only two structures are available of the free form of FVIIa with the first EGF-like domain resolved (6, 26). The first EGF-like domains, as represented in the structures of TF-bound and free FVIIa, cannot be overlaid without a clash between TF and the protease domain of FVIIa when oriented as in free FVIIa. An overlay of the protease domain structures does not cause a corresponding clash of the first EGF-like domain with TF. To further substantiate this interesting observation, we explored the configurational space around Gln-88. We found that the population of models which does not clash with TF is larger when the protease domain interacts with TF than when the light chain does (Figs. S1 and S2). This would indicate that

it is entropically favorable for the FVIIa protease domain to initiate complex formation with TF and be in agreement with the experimentally supported binding pathway (34, 35).

Of the two parts of the TF-interactive interface on FVIIa located on opposite sides of the linker region, the one involving the light chain contributes the majority of the binding energy (3). The virtually normal rate of dissociation of the established complex between sTF and des(83-86)- or des(84-86)-FVIIa, combined with the inability of sTF to stimulate the activity of these variants, thus suggests that the light chain is engaged in its native interactions with sTF with retained affinity and that the protease domain simultaneously interacts with sTF in an alternative, non-native mode which is capable of maintaining the normal affinity but not forming the allosterically important contacts between enzyme and cofactor. These crucial contacts are retained in the complexes between sTF and FVIIa variants lacking only one or two residues in the linker. The slow rate of association between des(83-86)- or des(84-86)-FVIIa and sTF suggests that a large proportion of the docking events mediated by the protease domain does not result in the formation of a stable complex because the suboptimal linker length forces the protease domain to slide relative to sTF to allow the light chain to bind. This is compatible with and a consequence of the mentioned preference of the FVIIa variants to exist in an energetically favorable conformation. At some stage during the transition process from encounter to stable complex, neither the initial nor the final interactions between FVIIa and sTF are established making the complex vulnerable and prone to dissociation. An alternative explanation for the binding kinetics (slow association and normal dissociation) of for instance des(84-86)-FVIIa could be that only a small fraction of the molecules are in a conformation which is at all compatible with sTF binding but do form the complex as we know it. However, this situation is less plausible due to the lack of detectable stimulation by sTF. Interestingly, the interaction of the isolated light chain of FVIIa with sTF, which like those of des(83-86)- and des(84-86)-FVIIa is of considerably lower affinity than that involving FVIIa, is primarily characterized by a much faster dissociation rate with a marginally affected association

rate (3). Altogether, the experimental and molecular modeling data obtained with the FVIIa variants with the shortest linkers support the hypothesis that FVIIa forms an encounter complex with sTF involving interactions between the protease domain and the cofactor, after which the light chain engages in binding to tether the proteins to each other. Our findings regarding FVIIa binding to TF are almost certainly pertinent also for the binding of FVII and for its ensuing activation. Enzyme and zymogen employ the same interactive interfaces on both sides of the linker in their very similar, perhaps identical, set of contacts with TF (36).

The linker region between the two EGF-like domains varies in length and sequence between the homologous factors VII(a), IX(a) and X(a). This may represent adaption to their respective cofactor (TF, factor VIIIa and factor Va, respectively) and serve as one of several players in the optimization of the intermolecular interplay. Factor IXa has the shortest linker and an extension has been shown to result in a diminished activity enhancement upon assembly with factor VIIIa (37). We have shown that the linker length is critical also in FVIIa and that a limited truncation results in a TF binding defect and lack of activity enhancement upon TF association. An extension will, at some point, most likely also become unfavorable and have functional consequences. Thus the linker length appears to be important for an optimal interaction, plausibly serving a spacer function between key cofactor-interactive interfaces on the enzyme surface. Moreover, data with factor Xa indicate that the linker sequence becomes more exposed and available for protein-protein interactions upon activation of the zymogen (38). This could be part of the explanation for the higher affinity of factor Va for factor Xa compared with X (39), definitely also involving rearrangements in the protease domain ensuing zymogen cleavage. An optimal combination of linker length and availability might be an important interaction determinant in this family of coagulation factors.

MATERIALS AND METHODS

Proteins and reagents

FVIIa, des(1-44)-FVIIa and sTF were prepared as previously described (11-13). Their concentrations were determined by absorbance measurements at 280 nm using absorption coefficients of 1.32, 1.26 and 1.5, respectively, for a 1-mg/ml solution and molecular masses of 50, 46 and 25 kDa, respectively. Biotin-E219C-sTF, a variant of sTF biotinylated on the introduced unpaired Cys residue, was a gift from Dr. Henrik Østergaard, Haemophilia Biochemistry, Novo Nordisk A/S. Human factor X (FX) was purchased from Enzyme Research Laboratories (South Bend, IN). The chromogenic substrates S-2288 and S-2765 were from Chromogenix (Milan, Italy), phospholipid-TGT from Rossix (Mölndal, Sweden), and potassium cyanate from Fluka.

Mutagenesis, expression and purification of FVIIa variants

Residues in FVII were deleted or inserted using the QuikChange kit (Stratagene, La Jolla, CA) and the human FVII expression plasmid pLN174 (14). The following sense (and complementary reverse) primers were used to generate the FVIIa linker variants: des86-FVIIa, 5'-CGGAACTGTGAGACGCACAAGGACCAGCTGATCTGTGTG-3'; des(85-86)-FVIIa, 5'-CGGAACTGTGAGACGCACGACCAGCTGATCTGTGTG-3'; des(84-86)-FVIIa, 5'-CGGAACTGTGAGACGGACCAGCTGATCTGTGTG-3'; des(83-86)-FVIIa, 5'-GGCCGGAACTGTGAGGACCAGCTGATCTGTGTGAACG-3'; 86G87-FVIIa, 5'-GAGACGCACAAGGATGGCGACCAGCTGATCTGTGTG-3'; 86GS87-FVIIa, 5'-GAGACGCACAAGGATGGCAGCGACCAGCTGATCTGTG-3'. The insertions between wild-type residues 86 and 87 in the two latter variants are underlined. Plasmids were prepared using a QIAfilter plasmid midi kit (Qiagen, Valencia, CA) and the introductions of the desired changes confirmed by sequencing. Baby hamster kidney cell transfection and selection, as well as protein expression, purification and autoactivation were performed as described (4, 15).

FVIIa activity assays

All assays were routinely performed at ambient temperature in 50 mM Hepes, pH 7.4, containing 0.1 M NaCl, 5 mM $CaCl_2$, 0.1% (w/v) PEG 8000 and 0.01% (v/v) Tween 80 (assay buffer). The amidolytic activity was measured using 300 nM FVIIa, des(1-44)-FVIIa or FVIIa linker variants in assay buffer. When studying the Ca^{2+} dependence of the activity, 2 mM EDTA replaced $CaCl_2$ in the assay buffer. Either fixed (1 mM) or varying (0.2-12.8 mM) concentration of S-2288 was used. The amidolytic activity dependence on ionic strength and Na^+ was also measured using 1 mM S-2288 in 50 mM Hepes, pH 7.4, containing 5 mM $CaCl_2$ or 2 mM EDTA, 0.1% (w/v) PEG 8000 and 0.01% (v/v) Tween 80, by including NaCl or choline chloride at concentrations ranging from 0.1 to 0.6 M. The effect of the cofactor TF on the amidolytic activity was assessed by mixing 10 nM FVIIa variant, no or 2-200 nM sTF, and 1 mM S-2288. Those FVIIa variants that were not stimulated at an [sTF] of 200 nM were also tested at 2 and 5 µM sTF. For proteolytic activity assessment, 1 µM FX was incubated with 50 nM FVIIa variant in a total volume of 100 µl for 20 minutes, whereafter FX activation was terminated by the addition of 50 µl assay buffer containing 20 mM EDTA instead of $CaCl_2$. The formed FXa was measured by the addition of S-2765 substrate (final concentration 0.5 mM). Alternatively, FX was used at a concentration of 150 nM in the presence of 25 µM phospholipids. Enzymatic activities were continuously monitored for 20 min (FVIIa or FVIIa:sTF amidolytic activity) or 5 min (FXa amidolytic activity) using a SpectraMax 190 kinetic microplate reader (Molecular Devices, Sunnyvale, CA).

Carbamylation

The FVIIa variants alone (2 µM) or FVIIa variants (0.5 µM) in the presence of 10 µM sTF were incubated with 0.2 M potassium cyanate in 50 mM Hepes, pH 7.4, containing 0.1 M NaCl, 5 mM $CaCl_2$, 0.1% (w/v) PEG 8000 and 0.01% (v/v) Tween 80. At time zero and after 20 and 60 minutes, 20-µl samples were transferred to a well containing 160 µl buffer and the amidolytic FVIIa activity was measured after the addition of 20 µl S-2288 substrate (final concentration 1 mM).

Surface plasmon resonance analyses

Biotin-E219C-sTF was injected at a concentration of 150 nM (in the assay buffer described above) and a flow rate of 20 µl/min for 100 s over a streptavidin (SA) chip (GE Healthcare Bio-Sciences AB, Uppsala, Sweden), resulting in an immobilization level of 90-120 resonance units. FVIIa, des86-FVIIa, des(85-86)-FVIIa and 86GS87-FVIIa were injected in two-fold dilutions in assay buffer from 64 to 4 nM, whereas des(84-86)- and des(83-86)-FVIIa were injected in 2-fold dilutions from 640 to 40 nM, all at a flow rate of 20 µl/min. Association and dissociation phases lasted for 4 and 10 minutes, respectively. Regeneration between runs was done with a 30-s pulse of 50 mM EDTA, pH 7.4, at 20 µl/min. All experiments were conducted using a Biacore 3000 instrument and evaluated using a 1:1 model in Biaevaluation 4.1 supplied by the manufacturer.

Molecular dynamics simulations

Free FVIIa and variants thereof with a deleted linker region between the EGF-like domains were analyzed by molecular dynamics. The models were based on the crystal structure of FVIIa in the FVIIa-TF complex (2, PDB entry code 1dan) after removal of the co-crystallized, covalently bound active-site inhibitor and TF. In FVIIa, the last 11 amino acid residues of the light chain are missing, but no attempts have been made to model the missing part to avoid perturbation of the X-ray data. The truncated structures were generated by deleting linker residues and rebuilding the remaining linker region. A final relaxation was conducted using 250 steps of conjugated gradient (CONJ) energy minimization in CHARMm (Accelrys Inc., San Diego, CA) using the CHARMM27 force field (16).

Molecular dynamics simulations were then performed using the program NAMD 2.5 (17) and the CHARMM27 force field for protein. The 9 calcium ions embedded in the X-ray crystal structure were preserved. The proteins were solvated (with TIP3P water molecules) in a periodic, truncated octahedron with box boundaries at least 6 Å from any given protein atom, and water molecules

were treated as rigid. Sodium and chloride ions were added to obtain an ionic strength of approximately 0.15. The resulting systems were composed of ~16,000 water molecules, ~ 42 sodium ions, ~ 43 chloride ions, and 9 calcium ions (in total approximately 54,000 atoms). A cut-off of 12 Å (switching function starting at 10 Å) for van der Waals interactions was applied. The particle mesh Ewald (PME) method was used to compute long-range electrostatic forces in all simulations (18). An integration time step of 1 fs was used. Langevin dynamics was applied to enforce constant temperature (T = 300 K) conditions. The Langevin dampening coefficient was set to 5 ps^{-1}. The pressure was maintained at 1 atm using the hybrid Nosé-Hoover Langevin piston method with a decay period of 100 ps and a dampening time constant of 50 ps (19, 20). Each system was equilibrated in the constant number, pressure and temperature (NpT) ensemble and simulated for at least 100 ns. The secondary structure of the linker was assigned using STRIDE (21) and the corresponding logos were generated by WebLogo (22).

Molecular modeling

The X-ray crystallographic structure of the FVIIa-TF complex (2, PDB entry code 1dan) was subject to structural modeling using the Rosetta modeling suite (23). It was processed by a Rosetta high-resolution protein structure refinement protocol consisting of three steps. First, stripping the structure of non-protein matter and patching the γ-carboxylated glutamic acid residues, then idealization of the geometry by IdealizeMover, and, finally, structural refinement with backbone perturbations and sidechain repacking by FastRelaxMover. This protocol was set up via the RosettaScripts interface (24). Using this protocol, 1,000 decoy models were generated for the wild-type FVIIa-TF complex and each of the complexes between TF and FVIIa with a truncated linker (des86-, des(85-86)-, des(84-86)-, and des(83-86)-FVIIa, respectively). The decoy models which scored best (lowest 10% evaluated by the Rosetta score 12 energy function) were chosen as the representative ensemble of the most probable structures. In this way, possible linker conformations were sampled while allowing for some flexibility in other parts of the structure including, of

particular interest, the interface between FVIIa and TF. Finally, the generated decoy complexes were structurally analyzed with attention on the FVIIa-TF interface (using InterfaceAnalyzer (25)) and the linker residues in FVIIa (the sequence flanked by the last Cys residue of the first EGF-like domain and the first Cys residue of the second EGF-like domain).

The configurational (φ,ψ)-space of backbone dihedral angles around the hinge residue Gln-88 was sampled in the X-ray crystallographic structure of FVIIa-TF (2) to quantify the available freedom of motion in FVIIa not causing a clash with TF. Two scenarios were investigated; one with the FVIIa protease domain aligned with TF as in the crystal structure and the FVIIa light chain rotated (roaming), and another with the FVIIa light chain aligned with TF and the protease domain roaming. The (φ,ψ)-space was exhaustively searched using 2 degree increments and all 32,400 generated decoys were scored without any further processing by the full-atom force field in Rosetta (score12).

ACKNOWLEDGMENTS

We thank Anette Østergaard for excellent technical assistance and Dr. Henrik Østergaard, Novo Nordisk A/S, for the gift of biotin-E219C-sTF.

REFERENCES

1. Davie EW, Fujikawa K, Kisiel W (1991) The coagulation cascade. Initiation, maintenance, and regulation. Biochemistry 30:10363-10370.
2. Banner DW, D'Arcy A, Chène C, Winkler FK, Guha A, Konigsberg WH, Nemerson Y, Kirchhofer D (1996) The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. Nature 380:41-46.

3. Persson E (1997) Characterization of the interaction between the light chain of factor VIIa and tissue factor. FEBS Lett 413:359-363
4. Persson E, Nielsen LS, Olsen OH (2001) Substitution of aspartic acid for methionine-306 in factor VIIa abolishes the allosteric linkage between the active site and the binding interface with tissue factor. Biochemistry 40:3251-3256
5. Furie B, Furie BC (1988) The molecular basis of blood coagulation. Cell 53:505-518.
6. Pike ACW, Brzozowski AM, Roberts SM, Olsen OH, Persson E (1999) Structure of human factor VIIa and its implications for the triggering of blood coagulation. Proc Natl Acad Sci USA 96:8925-8930.
7. Hagen FS, Gray CL, O'Hara P, Grant FJ, Saari GC, Woodbury RG, Hart CE, Insley M, Kisiel W, Kurachi K, Davie EW (1986) Characterization of a cDNA coding for human factor VII. Proc Natl Acad Sci USA 83:2412-2416.
8. Idusogie E, Rosen E, Geng J-P, Carmeliet P, Collen D, Castellino FJ (1996) Characterization of a cDNA encoding murine coagulation factor VII. Thromb Res 75:481-487.
9. Kurachi K, Davie EW (1982) Isolation and characterization of a cDNA coding for human factor IX. Proc Natl Acad Sci USA 79:6461-6464.
10. Leytus S, Chung DW, Kisiel W, Kurachi K, Davie EW (1984) Characterization of a cDNA coding for human factor X. Proc Natl Acad Sci USA 81:3699-3702.
11. Thim L, Bjoern S, Christensen M, Nicolaisen EM, Lund-Hansen T, Pedersen A, Hedner U (1988) Amino acid sequence and posttranslational modifications of human factor VIIa from plasma and transfected baby hamster kidney cells. Biochemistry 27:7785-7793.
12. Nicolaisen EM, Petersen LC, Thim L, Jacobsen JK, Christensen M, Hedner U (1992) Generation of Gla-domainless FVIIa by cathepsin G-mediated cleavage. FEBS Lett 306:157-160.

13. Freskgård P-O, Olsen OH, Persson E (1996) Structural changes in factor VIIa induced by Ca^{2+} and tissue factor studied using circular dichroism spectroscopy. Protein Sci 5:1531-1540.
14. Persson E, Nielsen LS (1996) Site-directed mutagenesis but not γ-carboxylation of Glu-35 in factor VIIa affects the association with tissue factor. FEBS Lett 385:241-243.
15. Persson E, Kjalke M, Olsen OH (2001) Rational design of coagulation factor VIIa variants with substantially increased intrinsic activity. Proc Natl Acad Sci USA 98:13583-13588.
16. McKerell Jr AD, Bashford D, Bellott M, Dunbrack Jr RL, Evanseck J, Field MJ, Fisher S, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher IWE, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-hydrogen empirical potential for molecular modelling and dynamics studies of proteins using the CHARMM22 force field. J Phys Chem 102:3586-3616.
17. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comp Chem 26:1781-1802.
18. Essman U, Perera L, Berkowitz T, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh method. J Chem Phys 103:8577-8593.
19. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. J Chem Phys 101:4177-4189.
20. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. J Chem Phys 103:4613-4621.
21. Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. Proteins 23:566-579.
22. Crooks G, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188-1190.

23. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman KW, Renfrew D, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban Y-E A, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P (2011) ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol 487:545–574.
24. Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, Ashworth J, Murphy P, Richter F, Lemmon G, Meiler J, Baker D (2011) RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. PLoS ONE 6:e20161.
25. Lewis SM, Kuhlman BA (2011) Anchored design of protein-protein interfaces. PLoS ONE 6:e20872.
26. Dennis MS, Eigenbrot C, Skelton NJ, Ultsch MH, Santell L, Dwyer MA, O'Connell MP, Lazarus RA (2000) Peptide exosite inhibitors of factor VIIa as anticoagulants. Nature 404:465-470.
27. Ohkubo YZ, Morrissey JH, Tajkhorshid E (2010) Dynamical view of membrane binding and complex formation of human factor VIIa and tissue factor. J Thromb Haemost 8:1044-1053.
28. Higashi S, Nishimura H, Aita K, Iwanaga S (1994) Identification of regions of bovine factor VII essential for binding to tissue factor. J Biol Chem 269:18891-18898.
29. Neuenschwander PF, Branam DE, Morrissey JH (1993) Importance of substrate composition, pH and other variables on tissue factor enhancement of factor VIIa activity. Thromb Haemost 70:970-977.
30. Neuenschwander PF, Morrissey JH (1994) Roles of the membrane-interactive regions of factor VIIa and tissue factor. J Biol Chem 269:8007-8013.
31. Schmidt AE, Shah P, Gauthier EM, Bajaj SP (2004) Protease domain sodium, calcium, and zinc sites in factor VIIa: crystal structures and kinetic studies. Blood 104:122a.

32. Bajaj SP, Schmidt AE, Agah S, Bajaj MS, Padmanabhan K (2006) High resolution structures of *p*-aminobenzamidine- and benzamidine-VIIa/soluble tissue factor. J Biol Chem 281:24873-24888.
33. Mosbæk CR, Nolan D, Persson E, Svergun DI, Bukrinsky JT, Vestergaard B (2010) Extensive small-angle X-ray scattering studies of blood coagulation factor VIIa reveal interdomain flexibility. Biochemistry 49:9739-9745.
34. Österlund M, Persson E, Svensson M, Carlsson U, Freskgård P-O (2005) Transition state analysis of the complex between coagulation factor VIIa and tissue factor: suggesting a sequential domain-binding pathway. Biochem Biophys Res Commun 327:789-793.
35. Österlund M, Persson E, Carlsson U, Freskgård P-O, Svensson M (2005) Sequential coagulation factor VIIa domain binding to tissue factor. Biochem Biophys Res Commun 337:1276-1282.
36. Kelley RF, Yang J, Eigenbrot C, Moran P, Peek M, Lipari MT, Kirchhofer D. (2004) Similar molecular interactions of factor VII and factor VIIa with the tissue factor region that allosterically regulates enzyme activity. Biochemistry 43:1223-1229.
37. Celie PHN, van Stempvoort G, Fribourg C, Schurgers LJ, Lenting PJ, Mertens K (2002) The connecting segment between both epidermal growth factor-like domains in blood coagulation factor IX contributes to stimulation by factor VIIIa and its isolated A2 domain. J Biol Chem 277:20214-20220.
38. Ambrosini G, Plescia J, Chu KC, High KA, Altieri DC (1997) Activation-dependent exposure of the inter-EGF sequence Leu^{83}-Leu^{88} in factor Xa mediates ligand binding to effector cell protease receptor-1. J Biol Chem 272:8340-8345.
39. Persson E, Hogg PJ, Stenflo J (1993) Effects of Ca^{2+} binding on the protease module of factor Xa and its interaction with factor Va. J Biol Chem 268:22531-22539.

Table I. Kinetic parameters of the binding of FVIIa and FVIIa linker variants to sTF measured by surface plasmon resonance. Biotinylated sTF was captured on a streptavidin-coated chip and the binding of serial dilutions of the FVIIa variants was measured. The kinetic parameters (mean ± SEM) were determined using the Biacore 3000 evaluation software (Biaevaluation 4.1) and a model based on 1:1 binding stoichiometry.

Ligand	k_{on} (x $10^5 M^{-1}s^{-1}$)	k_{off} (x $10^{-3}s^{-1}$)	K_d (nM)
86GS87-FVIIa	3.7±0.3	1.2±0.1	3.3±0.3
FVIIa	4.5±0.3	1.3±0.2	3.0±0.4
des86-FVIIa	3.7±0.2	1.2±0.0	3.3±0.1
des(85-86)-FVIIa	2.7±0.4	1.8±0.1	6.8±0.6
des(84-86)-FVIIa	0.054±0.003	1.7±0.2	310±30
des(83-86)-FVIIa	0.052±0.003	1.4±0.0	270±10

FIGURE LEGENDS

Figure 1. Structures and secondary structure assignments of the linker region in free FVIIa variants based on molecular dynamics simulations. *Top,* representative structures (30-ns snapshots) showing the distance in Å (illustrated by a grey line) between the C_α atoms of Cys-81 and Cys-91 (full-length numbering). The backbone traces start at residue Phe-71 in the first EGF-like domain and the sidechains are shown for all residues present in the linker region and the flanking Cys residues. *Bottom,* secondary structure assignments of the linker region based on the simulations shown as weblogos (22). The one-letter secondary structure codes are: H, α-helix; G, 3_{10}-helix; I, PI-helix; E, extended conformation (sheet); B, isolated bridge; T, turn; C, random coil (none of the previous).

Figure 2. Functional response of FVIIa linker variants upon sTF association. *A,* 10 nM FVIIa (green circles), des86-FVIIa (□), des(85-86)-FVIIa (■), des(84-86)-FVIIa (○) or des(83-86)-FVIIa (●) was incubated with the indicated concentrations of sTF and 1 mM S-2288. The ratio of the amidolytic activities of a FVIIa variant in the presence and absence of sTF is shown as a function of [sTF]. The results obtained with des(83-86)- and des(84-86)-FVIIa are virtually identical and the data points representing des(84-86)-FVIIa (the open circles) are therefore not visible. *B,* the corresponding data shown for 86G87-FVIIa (△) and 86GS87-FVIIa (▲) with the FVIIa data from *A* included for comparison. The data shown are from a representative experiment (n = 4).

Figure 3. Kinetics of the binding of FVIIa linker variants to sTF measured by surface plasmon resonance. Corrected sensorgrams are shown for the interactions between sTF and des(84-86)-FVIIa (A), des(85-86)-FVIIa (B), FVIIa (C), and 86GS87-FVIIa (D). Des(84-86)-FVIIa was injected in a 2-fold dilution series in concentrations from 40-640 nM (represented by the curves from bottom to top), whereas the other three variants were injected in a 2-fold dilution series

in concentrations from 4-64 nM. Experimental data are in black and fitted data in red. Binding curves for des(83-86)-FVIIa and des86-FVIIa are not shown here, but the derived constants are included in Table 1.

Figure 4. Molecular modeling of FVIIa-TF complexes. Average scores and standard deviations in negative Rosetta energy units (REUs) of the best scoring 10% of the models of complexes between FVIIa linker variants and TF are shown. The Ramachandran plots visualize the backbone dihedral angles (phi, psi) of the amino acid residues in the linker region in these complexes. *wt*, wild-type FVIIa; *desX-FVIIa*, FVIIa variant with residue(s) X deleted.

Figure 5. Residual amidolytic activity of the apo (Ca^{2+}-free) forms of FVIIa linker variants. The residual activity after the removal of Ca^{2+} from FVIIa, FVIIa linker variants and des(1-44)-FVIIa was measured in the presence of 0.1 M NaCl (*black bars*) or choline chloride (*red bars*). All data are mean ± S.D., $n = 3$.

Figure S1. Contour-plot of the clash score (in Rosetta Energy Units) upon perturbation of Gln-88 backbone dihedral angles when the LC of FVIIa is "roaming". The contour level corresponding to a reasonable cut off clash score (zero) is drawn with thick lines. The circle (orange) shows the native dihedral angles observed in the crystal structure.

Figure S2. Contour-plot of the clash score (in Rosetta Energy Units) upon perturbation of Gln-88 backbone dihedral angles when the HC of FVIIa is "roaming". The contour level corresponding to a reasonable cut off clash score (zero) is drawn with thick lines. The circle (orange) shows the native dihedral angles observed in the crystal structure.

FIGURES

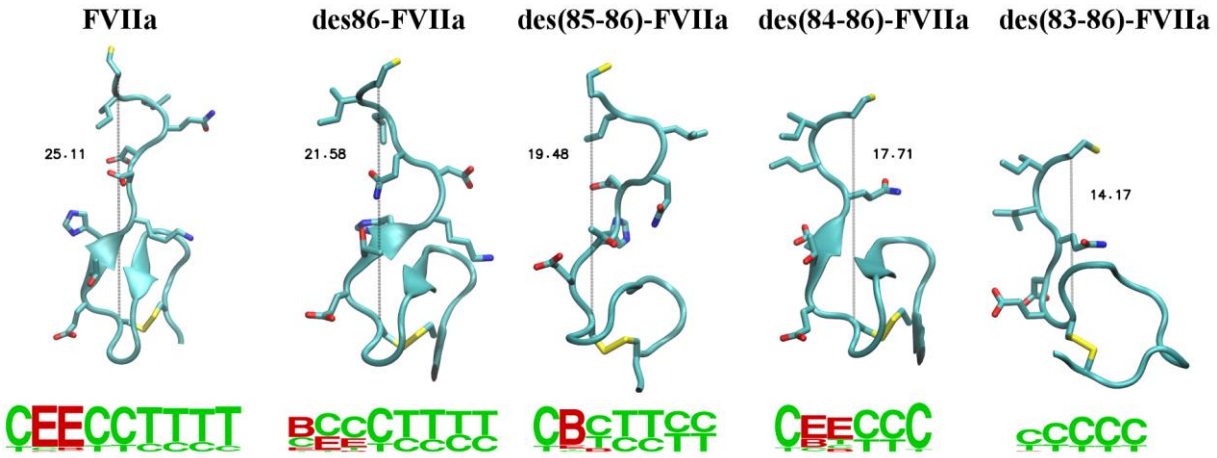


Figure 1

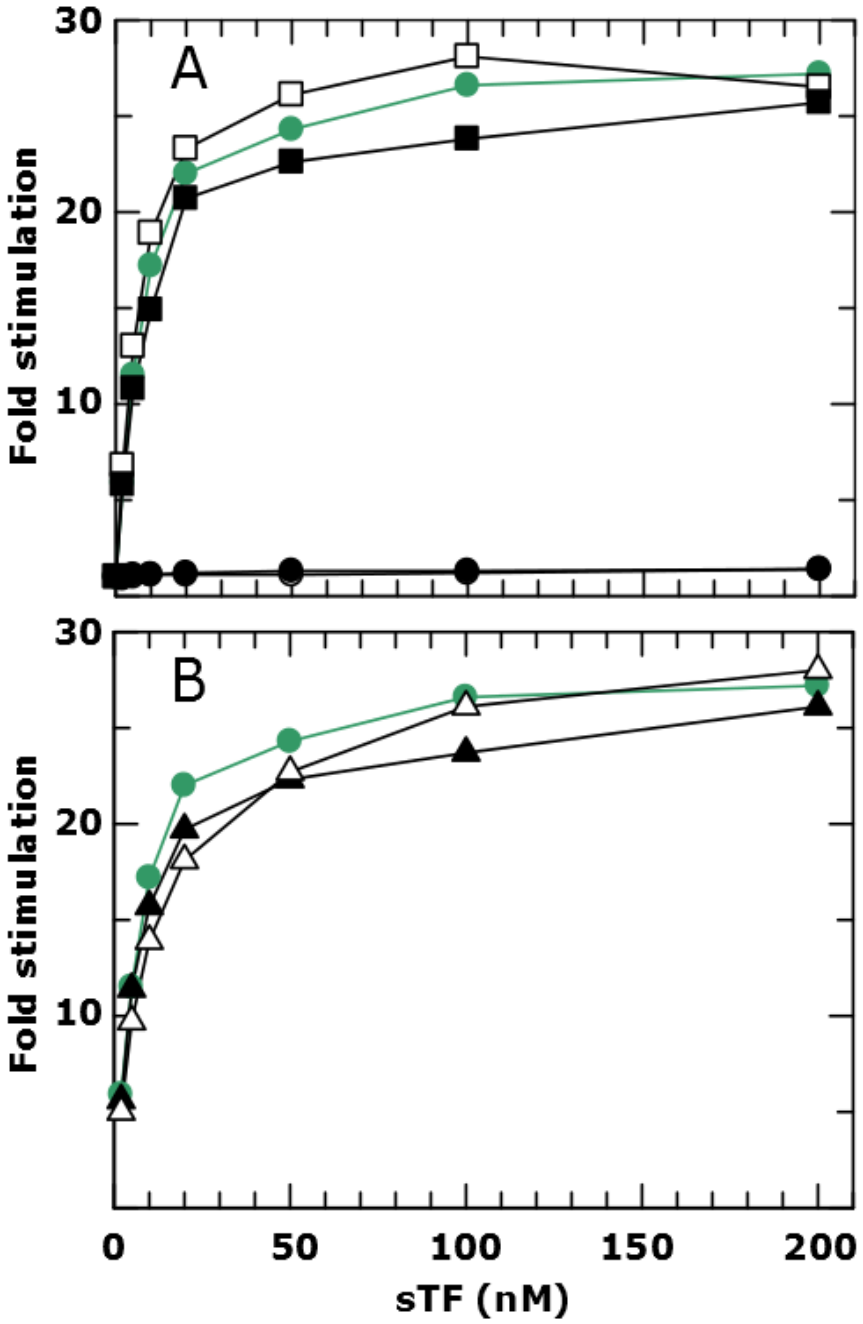

A
B
Fold stimulation
Fold stimulation
30
20
10
30
20
10
0
50
100
150
200
sTF (nM)

Figure 2

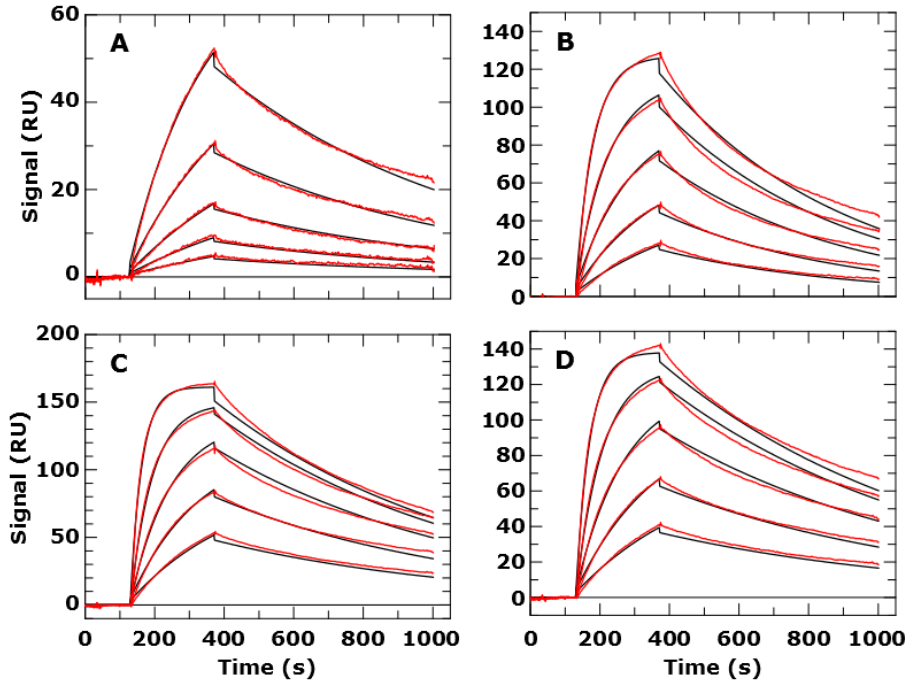

A
B
C
D
Signal (RU)
Signal (RU)
Time (s)
Time (s)
0
20
40
60
0
20
40
60
80
100
120
140
0
50
100
150
200
0
20
40
60
80
100
120
140
0
200
400
600
800
1000

Figure 3

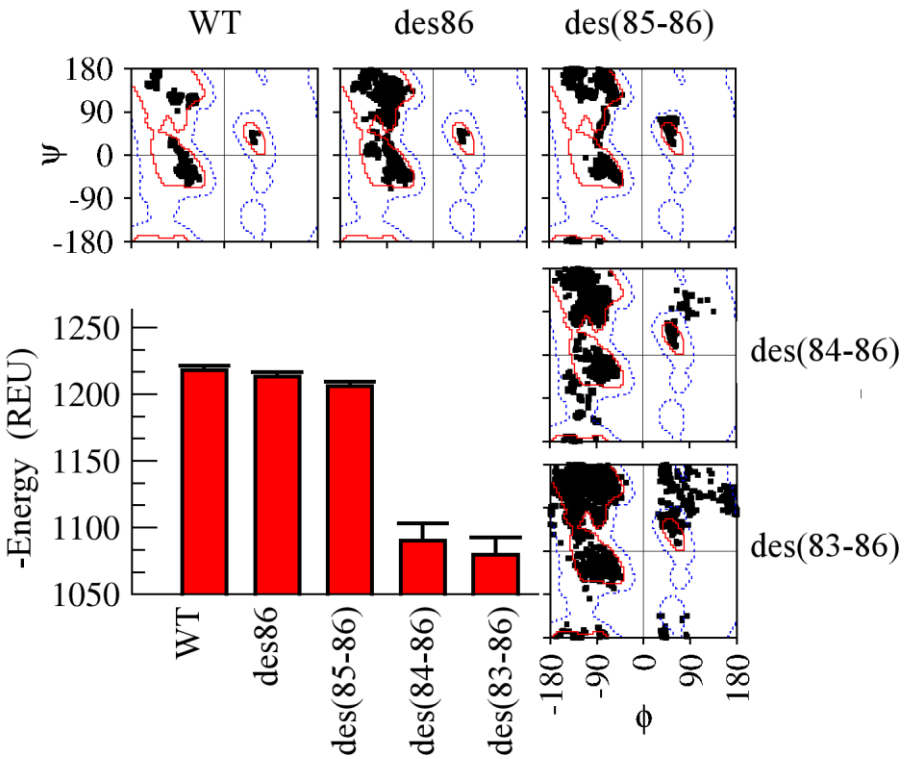

WT
des86
des(85-86)
des(84-86)
des(83-86)
180
90
0
-90
-180
ψ
-Energy (REU)
1250
1200
1150
1100
1050
WT
des86
des(85-86)
des(84-86)
des(83-86)
-180
-90
0
90
180
ϕ

Figure 4

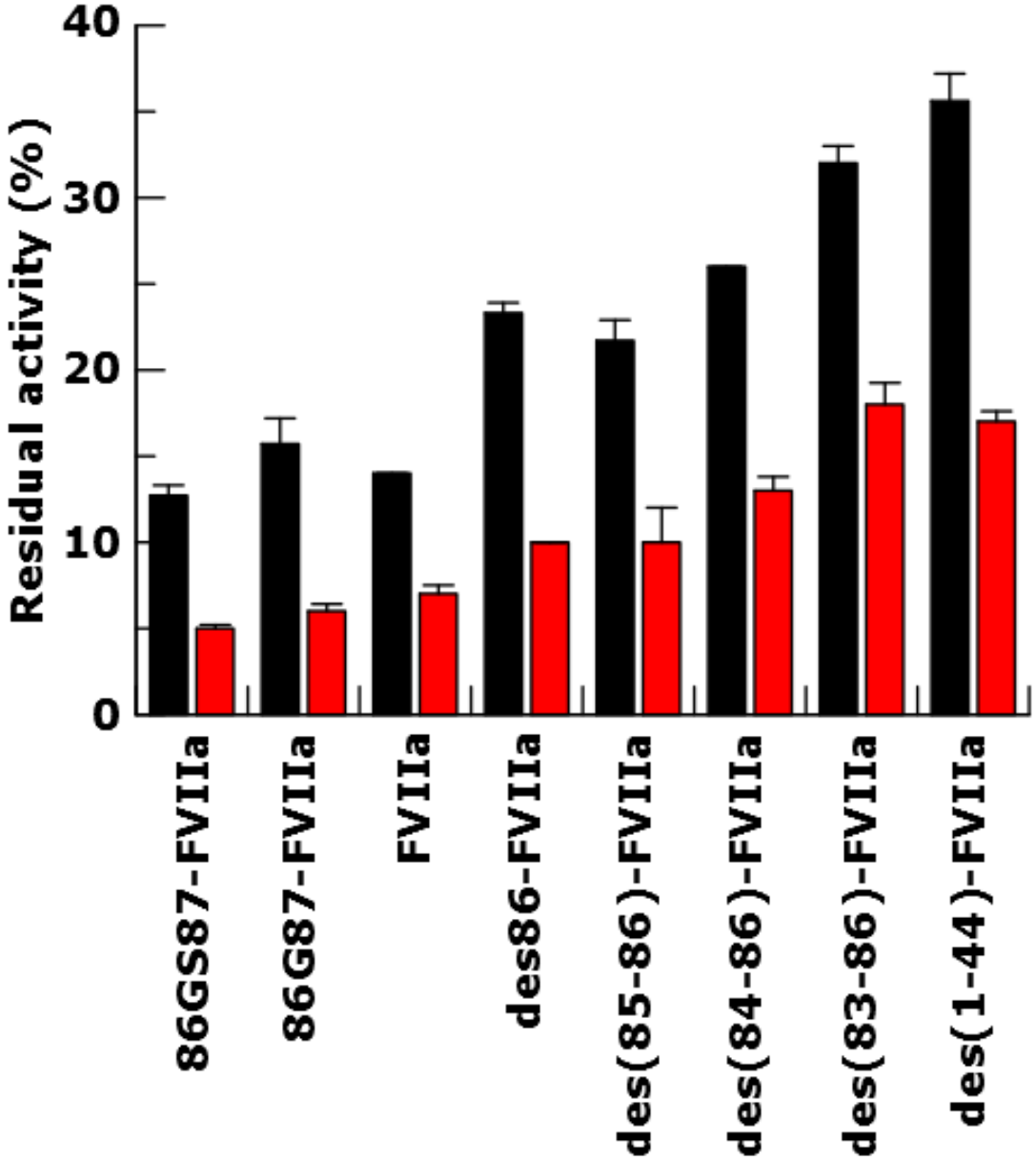

Residual activity (%)
40
30
20
10
0
86GS87-FVIIa
86G87-FVIIa
FVIIa
des86-FVIIa
des(85-86)-FVIIa
des(84-86)-FVIIa
des(83-86)-FVIIa
des(1-44)-FVIIa

Figure 5

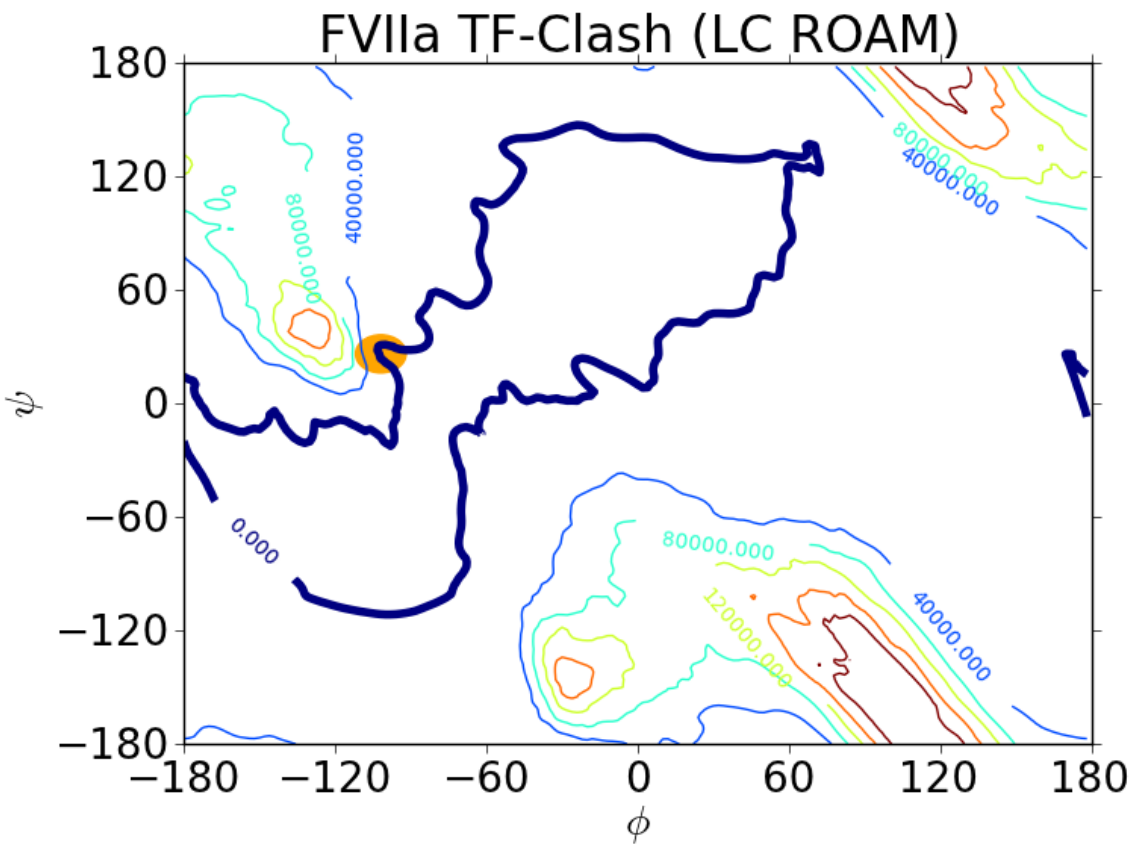

FVIIa TF-Clash (LC ROAM)
180
120
60
0
−60
−120
−180
−180
−120
−60
0
60
120
180
ψ
ϕ
0.000
40000.000
80000.000
120000.000

Figure S1

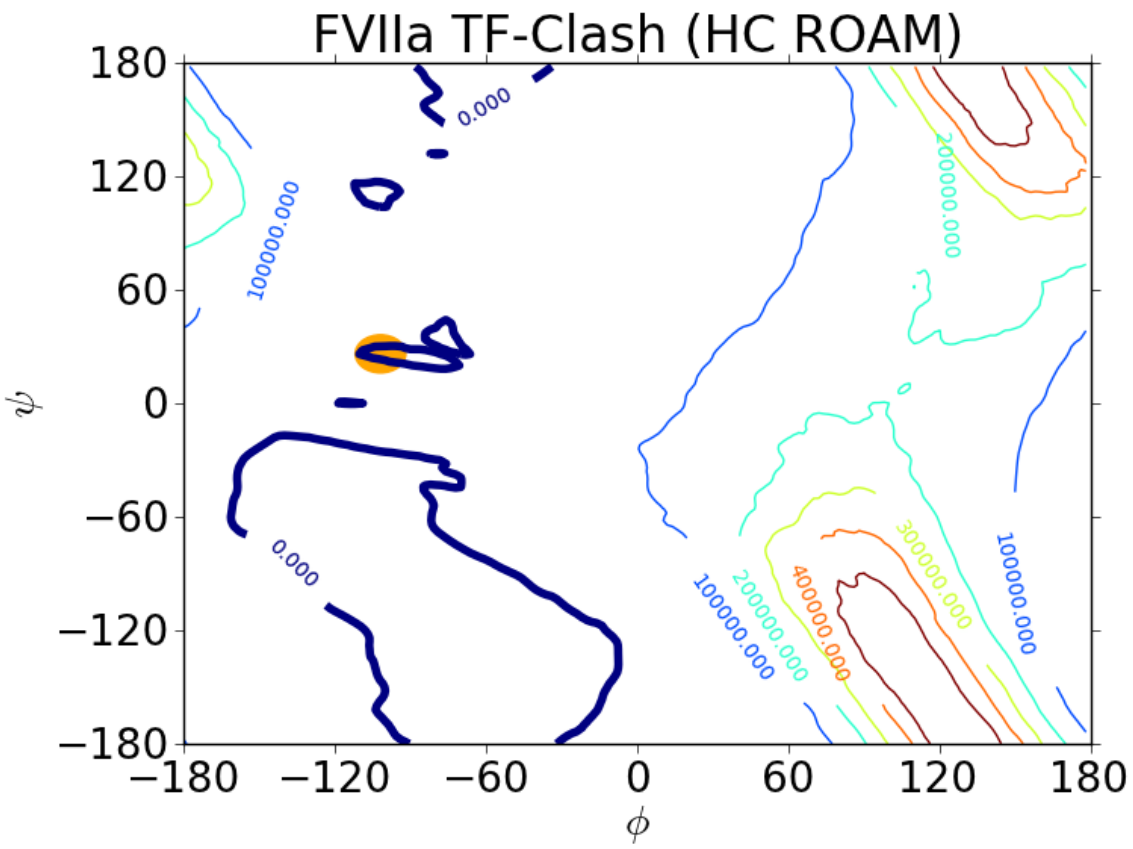

FVIIa TF-Clash (HC ROAM)
180
120
60
0
−60
−120
−180
−180
−120
−60
0
60
120
180
ψ
ϕ
0.000
100000.000
200000.000
300000.000
400000.000
100000.000
200000.000
100000.000
0.000

Figure S2

4. Intrinsic tenase complex

4.1 Trauma-site directed cofactor function of FVIIIa

4.1.1 Prelude

Manuscript 4 described a study of the membrane-interaction of the two C-terminally positioned tandem discoidin domains of FVIIIa, FVIII C1 and FVIII C2. Parallel simulation trajectories of spontaneous membrane-binding of achieved for each domain through the use of a recent membrane-mimetic model(Y Zenmei Ohkubo, Pogorelov, Arcario, Christensen, & Tajkhorshid, 2012). The FVIII C1 and FVIII C2 share ~40 % sequence identity and the domains cut-out from the crystal structure of FVIII (PDB entry code 3cdz(Ngo, Huang, Roth, Furie, & Furie, 2008)) structurally align with sub-Ångstrøm RMS. Considering these similarities, our results are unexpected and surprising.

The actual mechanism of protein binding to membranes is often complex because it is a combination of several important components. One important aspect in this context is the hydrophobic partitioning where parts of the protein enter the membrane to be shielded from water. This (partial) insertion of hydrophobic residues or loops follows the initial attraction and contact with the membrane, which is governed by electrostatics, likely the most important long-range interaction leading to protein-membrane binding.(Bardelle, Furie, Furie, & Gilbert, 1993) At the continuum level of theory, Poisson-Boltzmann theory can be used to understand the interfacial region between the membrane and water. However, inherent limitations of such descriptions does not account for local variation via e.g. membrane-composition and specific interactions that can lead to accumulation of lipids near the binding region. Furthermore, the conformation of the protein plays a critical role for the membrane-binding. In the case of FVIII, intriguingly, the non-activated and the activated molecules exhibit distinct binding-dependencies of the PS component of the membrane even though there are no cleavage sites within the membrane-anchoring modules (FVIII C1 and FVIII C2).(Engelke, Lippok, Dorn, Netz, & Rädler, 2011)

Our simulation results reveal that both FVIII C2 and FVIII C1 insert peripherally via extruding β-hairpin loops (the spikes, S1-S4), but whereas FVIII C2 stands almost normal

to the membrane plane, FVIII C1 can adopt several moderately to highly tilted orientations, possibly enabling the full-length cofactor (FVIII) to accommodate an appropriate interaction in the tenase complex, FVIIIa:FIXa, and in the ternary complex FVIIIa:FIXa:FX.

4.1.2 Manuscript 4: Madsen JM, *et al.*, submitted

Membrane-interaction of the factor VIIIa discoidin domains: implications for the tenase complex

Jesper J. Madsen[1,2], Y. Zenmei Ohkubo[3], Emad Tajkhorshid[3,4], Günther H. Peters[2], Johan H. Faber[5], Ole H. Olsen[1]

[1] Haemophilia Biochemistry, and [5] Protein Characterisation, Novo Nordisk A/S, DK-2760 Måløv, Denmark.
[2] Department of Chemistry, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.
[3] Beckman Institute for Advanced Science and Technology, and [4] Center for Biophysics and Computational Biology, School of Chemical Sciences, Departments of Biochemistry and Chemistry, College of Medicine, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

Correspondence: Ole H. Olsen, Haemophilia Biochemistry, Novo Nordisk A/S, Novo Nordisk Park, F5.1.29, DK-2760 Måløv, Denmark.
Tel.: +45 3075 4511
E-mail: oho@novonordisk.com

Key phrases Blood Clotting; Coagulation Factor VIII; fXase; intrinsic tenase; Membrane Proteins; Molecular Dynamics Simulations

Running title FVIII C1/C2 interaction with membranes

Abstract

A recently developed membrane-mimetic model was applied to study the membrane-interaction of the two anchoring C2-like discoidin domains of human coagulation factor (F)VIIIa. Both individual domains, FVIII C1 and FVIII C2, were observed to enter the phospholipid membrane unguided by partial or full insertion of the extruding loops (the spikes). FVIII C1 and FVIII C2 adopted different molecular orientations in the membrane-bound state; FVIII C2 roughly positioned normal to the membrane plane while FVIII C1 displayed a multitude of tilted orientations. The results indicate that FVIII C1 may be important in modulating the orientation of the FVIIIa molecule to accommodate the interaction with FIXa, which is anchored to the membrane via its γ-carboxyglutamic acid-rich (Gla)-domain. Additionally, a structural change was observed in FVIII C1 in the coiled main chain leading into the first spike. A tight interaction with one lipid per domain, similarly to what have been suggested for the homologous FVa C2, is characterized. Finally, we rationalize known FVIII antibody epitopes and the scarcity of documented hemophilic missense mutations related to improper membrane binding of FVIIIa based on the prevalent non-specificity of ionic interactions in the simulated membrane-bound states of FVIII C1 and FVIII C2.

Introduction

Using theoretical and computational modeling, a detailed characterization of the individual C2-like discoidin domains from human coagulation factor (F)VIIIa in the membrane-bound state with respect to lipid specificity, molecular orientation and flexibility is presented. These are properties of fundamental importance for the cofactor activity of FVIIIa in the intrinsic tenase (FX activating) complex. The molecular dynamics (MD) simulation technique provides spatial and temporal resolution fine enough for generating detailed information on the atomic level of the membrane binding and lipid interaction of the FVIIIa C2-like domains. In order to expedite the dynamics of membrane lipids, the use of a Highly Mobile Membrane-mimetic Model (HMMM)(1) is adopted. This enables the generation of multiple independent parallel trajectories and hence, a statistical representation of the dynamics becomes feasible.

The biochemical processes governing blood clotting is classically represented by a 'waterfall' or 'cascade' model(2)(3). Two distinct pathways (intrinsic vs extrinsic) funnel into a common amplification phase where activated FX (FXa) is generated which, together with its cofactor FVa, is responsible for the burst of thrombin that subsequently leads to fibrin clot formation. The biomolecular components FVIII and FIX are circulating the bloodstream as the inactive precursors that, upon activation, assemble on a phospholipid surface into the highly potent tenase complex. Hemophilia A is characterised by deficiency of FVIII observed as low levels or dysfunction of the protein procofactor or the presence of inhibitory antibodies and is by far the most common bleeding disease.(4)

A pivotal aspect of the coagulation cascade is the ability of restricting blood clotting to the injury site. The platform for this spatial localization is provided by the activated platelet membrane surfaces, which attract and stimulate activity by means of both membrane composition and the presence of elevated levels of certain cofactors to the coagulation enzymes. The tenase components FVIIIa and FIXa have the ability to selectively recognize this platform, and once properly bound and the binary complex formed, catalytic efficiency of FIXa is up-regulated by approximately five orders of magnitude(5). Their membrane binding modes, however, are quite different; FIXa is anchored to the membrane by its vitamin K-dependent γ-carboxyglutamic acid-rich (Gla)-domain while the membrane-

targeting modules of FVIIIa are the two C2-like discoidin domains, C1 and C2, which recognize phosphatidylserine (PS)-containing platelet or endothelial cell membranes in a Ca^{2+}-independent manner(6). Since either domain alone, FVIII C1 or FVIII C2, by itself seemingly has the ability to recruit the entire cofactor molecule to phospholipid membranes(7), optimal biological activity most certainly requires both.

The active cofactor molecule, FVIIIa, is a trimer and consists of five major domains (A1, A2 and the light chain A3-C1-C2) with a total of >1,200 amino acid residues. The structural topology of the homologous C2-like domains is that of lectin and commonly known as a jelly-roll β-barrel (Fig. 1A); eight anti-parallel β-strands are arranged in two major β-sheets, wrapped to form the barrel and then flattened to a sandwich-like shape.(8) Connecting the β-strands at the bottom of the barrel are four hairpin loops also called the spikes (S1-S4) or fatty feet(9), the latter designation being due to the presence of multiple solvent exposed hydrophobic residues. These spikes are of particular importance for platelet membrane-aided functionalities because S1-S4 are hypothesized to insert into the hydrophobic core of the phospholipid membrane(10)(11). For this reason, much attention has been dedicated in the literature to elucidate how affinity and specificity of the FVIIIa molecule (and FVIII C2 on its own) toward phospholipid membranes is facilitated by the residues in these spikes (alanine mutagenesis(12), motif mutagenesis(13) and loop-swaps(9)). The primary membrane-anchoring domain of FVIIIa is conventionally thought to be the C2 domain. Recently, however, several studies have emphasized the important role of the C1 domain in the membrane-mediated cofactor function of FVIIIa.(14)(7)(15)(16)

Previous studies on the molecular orientation of FVIIIa suggest a crystal structure-like domain configuration with a membrane-bound configuration in which the molecule is close to perpendicular (or slightly tilted) and both FVIII C1 and FVIII C2 are peripherally inserted in the membrane.(17)(18)(19) A fundamentally different interaction, requiring large scale domain rearrangements to accommodate a binding mode where solely FVIII C2 interacts with the membrane and is deeply inserted, has been suggested based on the cryo-electron microscopy technique in conjugation with a membrane model consisting of lipids assembled on nanotubes.(20)(21)(22)

The article is structured as follows. The protein-lipid interaction of each of the two membrane-anchoring discoidin domains from human FVIIIa (FVIII C2 and FVIII C1) is characterized, and a PS-binding motif is described. The achieved membrane-bound states are discussed in relation to haemophilic disease mutations, antibody binding epitopes, and a putative tenase complex model.

Materials and Methods

Molecular Dynamics Simulations.

Structures and Setup of the HMMM.

The starting structures of the discoidin C2-like domains were taken from the crystallized B-domainless human FVIII with PDB entry code 3cdz(17) (residues 2021-2172 for FVIII C1; residues 2173-2332 for FVIII C2). It can be assumed that there is no structural distinction between said domains in FVIII and FVIIIa neither in the context of the full-length (pro)cofactor nor cut-out on their own. pKa calculations were done using PROPKA(23) to verify that the standard protonation states appropriately represent a pH value of 7; arginines and lysines are positively charged, aspartic and glutamic acids are negatively charged, and histidines are neutral.

To assemble the HMMM surface patch(1), short-tailed (st-)lipid molecules truncated at the C_5 carbon atom containing a PS headgroup were packed to sandwich a layer of 1,1-dichloroethane (DCLE) solvent molecules with the headgroups facing the water phase and acyl chains against the DCLE phase using Packmol(24).

The individual C2-like domains (Fig. 1A) were then hand placed such that the β-barrel central axis (the third principal axis of inertia) was oriented approximately perpendicular ($\theta =\sim 0$; Fig. 1B) to the membrane plane and with the hydrophobic membrane-anchoring spikes (S1 is residues 2043-2046 for FVIII C1 and 2196-2203 for FVIII C2; S2 is residues 2056-2059 for FVIII C1 and 2213-2217 for FVIII C2; S3 is residues 2089-2096 for FVIII C1 and 2248-2255 for FVIII C2; S4 is residues 2156-2159 for FVIII C1 and 2313-2316 for FVIII C2) of the domain facing the membrane (Fig. 1B). The resulting systems (denoted C2/HMMM and C1/HMMM) which contained an HMMM patch and a C2-like domain were solvated using the Solvate functionality in VMD(25) and neutralized with sodium (Na^+) ions. Water molecules placed by Solvate in the organic solvent phase of the membrane (due to relatively large gaps) were removed. A reference system was prepared in a similar fashion and contained only FVIII C2 in bulk water (C2/Soln) at physiological 0.15 M NaCl. Detailed overview and component count of the prepared systems can be seen in Table 1.

Simulation details.

The prepared systems (C2/HMMM and C1/HMMM) were treated in the following manner. The potential energy was initially minimized by applying the conjugated gradient method for 5,000 steps. Then the systems were subjected to a short (100 picosecond (ps)) simulation in the *NPT* ensemble with constant membrane plane (*x*/*y*) aspect ratio to resolve imperfect packing of DCLE molecules, and hence equilibrate the membrane thickness. During this phase, the area per lipid in the systems was free to change and resulting values ranged from 85.0 Å to 88.5 Å (prior to protein insertion), corresponding to reasonable values for fluid phase lipid bilayers(26). Once the thickness had adjusted, a mild harmonic constraint with force constant 0.01 kcal/mol was applied on the C_2 carbon atoms of all st-lipids to restrain them toward the average position of all C_2 atoms in their respective membrane leaflet. This was done to gently reduce vertical diffusion (along the *z*-axis) of the st-lipids and to discourage lipid inversion. After this point, the production MD simulations was performed 5 times independently of each system in the *NPzAT* ensemble with the target P = 1 atm and T = 310 K controlled by the Nosé-Hoover Langevin piston barostat(27)(28) and the Langevin thermostat (damping coefficient: 5/ps)(29), respectively. Throughout, pressure coupling was applied along the *z*-axis (piston damping coefficient: 5/ps, piston period: 100 femtosecond (fs), and piston decay: 50 fs). Long-range electrostatic forces were calculated using the Particle Mesh Ewald (PME) method(30)(31) with a grid spacing of approximately 1 Å and a fourth-order spline for interpolation. Electrostatic forces were updated every 4 fs. Van der Waals interactions were cut off at 12 Å in combination with a switching function beginning at 10 Å. Periodic boundary conditions were applied in *x*-, *y*-, and *z*-directions. All MD simulations were performed using the NAMD 2.9 software package(32) with the CHARMM36 force field(33), CMAP correction(34) and the TIP3P water model(35). An integration time step of 2.0 fs was used for the velocity Verlet algorithm with SHAKE(36). All analysis and visualization of molecular structures was done using the Visual Molecular Dynamics (VMD) 1.9.1 software package(25); calculation of van der Waals and electrostatic interaction energies between protein and the HMMM was done using the NAMDEnergy plugin without the PME. Plots were prepared using Grace (xmgrace, http://plasma-gate.weizmann.ac.il/Grace).

Calculation of Order Parameters for Backbone amide (C-N) vectors

The amide vector was defined for each individual amino acid residue as the normalized vector pointing from the backbone carbonyl carbon C (atom) to the backbone amide nitrogen N. The protein was aligned over the trajectory by a root mean square deviation-based structural alignment on heavy atoms. Following this, nematic order parameters were calculated as(37)

$$S_i^2 = \frac{3}{2}\langle \hat{z}_i . \hat{d}_i \rangle^2 - \frac{1}{2}$$

where S_i^2 is the order parameter for the *i*'th residue, $\hat{d}_i$ is a unit vector describing the trajectory-averaged C-N direction, and $\hat{z}_i$ is the instantaneous atomic unit vector. The brackets imply averaging over the trajectory. The frame rate for the analysis was 10 frames per ns. The order parameters assume values from 0 to 1 and they are a measure of how much the atomic amide vectors fluctuate around their respective trajectory-averaged directions, indicating the flexibility of the backbone at that given position in the structure.

Construction of a putative model of the FVIIIa:FIXa tenase complex

A putative model of the FVIIIa:FIXa tenase complex was constructed on the basis of the recently published X-ray crystallographic structure of the FVa:FXa prothrombinase complex from the venom of the eastern brown snake(38). The light chain of FIXa in the constructed tenase complex was modeled in a FVIIa-like extended conformation(39), instead of an arched conformation adopted by procine FIXa with a partially disordered Gla-domain(L chain of PDB entry code 1pfx(40)). Homologous molecules (FVa versus FVIIIa and FIXa versus FXa) were aligned and the building of missing loops was done with the molecular modeling package Quanta(Molecular Simulations Inc., San Diego, CA). The atomic coordinates of the model are available upon request to the authors.

Results and Discussion

The positively charged C2-like domain is attracted to the negatively charged phospholipid membrane due to favorable electrostatic interactions, shown by the calculated non-bonded interaction energies between the domain and the membrane (Figs. 2, *bottom* & S1, in the

Supporting Material). As we describe in the following, the five parallel FVIII C2 trajectories converged into a stable membrane-bound state with comparable spike insertion depths and domain tilt angles (Fig. 1C). Interestingly, while FVIII C1 carries a larger net positive charge, +9 *e* for FVIII C1 and +4 *e* for FVIII C2, respectively, it is FVIII C2 that is the stronger electric dipole with 450.1 D compared with 421.8 D (based on the X-ray crystallographic structure; differences are further accentuated during the course of the simulations, data not shown). The amino acid residues directly involved in the membrane-binding process and their specific sequence of interaction varied depending on the orientation of the studied FVIIIa domain upon membrane contact. In the membrane-bound state of FVIII C2, a snug PS-specific interaction via either R2220 or R2320 (Fig. 3) was achieved. The same result was not observed for the four membrane binding FVIII C1 trajectories (one trajectory did not result in spontaneous binding and was discarded from further analysis) which instead displayed a broad range of domain tilt angles and significant structural modulations. Furthermore, the characteristic tight PS-specific interaction with the corresponding arginine R2320-counterpart in FVIII C1, R2163, was only established in a single trajectory. The observed tightly bound lipid is further stabilized by polar interactions of the amino and phosphate groups with residues of the protein, but these depended on the lipid orientation at the PS specificity pocket(10) counterpart in FVIII C1 or FVIII C2. These findings show that either domain can achieve a protein-lipid interaction that is specific to PS. However, the observed propensities are surprising considering experimental reports, which show that the FVIII C2 binding to vesicles is not specific to PS.(41)

FVIII C2 adopts a perpendicular membrane-bound configuration and achieves a PS-specific interaction.

The converged molecular orientation is characterized by the domain tilt angle (Fig. 1B), which exhibited shallow angles of 20-40 degrees (Fig. 4, *top*) and hence is consistent with the previously proposed perpendicular (or slightly tilted) mode of interaction(17)(18) and also entry in the Orientation of Proteins in Membranes (OPM) database(42) for FVIII (PDB entry code 2r7e(43)). Regions of FVIII C2 in contact with the membrane were largely confined to the spikes (Fig. 5, *top*). No significant structural modulations are observed for FVIII C2 as indicated by the calculated order parameters both for the

membrane-bound simulations, C2/HMMM, (Fig. 5, *top*) and for the FVIII C2 placed in bulk water, C2/Soln (Fig. S2), albeit a mild overall stabilization is observed and consistent with a recently published study of FVIII C2 by hydrogen-deuterium exchange mass spectrometry(44). The extruding hydrophobic spikes (Fig. 1A) are inserted peripherally into the membrane as the relative height of the domain tends to decreases over time until convergence is achieved (Fig. 2, *left*). The membrane-contacting surface is confined to about one third of the whole domain that mostly includes S1-S4 (Fig. 1C). The basic residues of this region of the FVIII C2 domain interact with the acidic PS lipids (either $<PO_4^->$ or $<COO^->$ functional group) in a mainly transient and non-specific way with two exceptions, R2220 and R2320. It is the unique combination of a soft funnel-like geometry of the loops in the membrane-interacting part of the domain and a centrally positioned arginine (either R2320 or R2220 depending on the domain tilt angle) that interacts selectively with the carboxylic acid group of the lipid that defines the PS-specific interaction (Fig. 3). This binding motif is further stabilized by polar interactions of the phosphate group of the lipid with surrounding amino acid residues of the protein. It is interesting to note that while PS-specific FVIII C2-lipid interactions do occur, occasionally, via both R2320 and R2200, they seem to be mutually exclusive; in domain orientations that enable R2220 to be selectively engaged with the $<COO^->$ group of a PS lipid, R2320 cannot (and vice versa). Furthermore, the PS-specific interactions are observed less frequent than for FVa C2 studied previously(45). In the present study, for FVIII C2, the key interaction of R2220 is found in two of the five trajectories (#3 and #4) and present in about ~80 % of the time in these two cases. The R2320 key interaction is also found in two other trajectories, #5 and #3, but only ~33 % and ~10 % of the time, respectively. A direct comparison with FVa C2 is complicated by subtle yet important differences in amino acid sequences and, in particular, by the fact that FVa C2 carries a higher net positive charge than FVIII C2 (as determined by a surplus of basic K and R residues over acidic D and E residues of +12 *e* for FV C2 and +6 *e* for FVIII C2, respectively), differences that necessarily affect membrane affinity and association kinetics. While binding free energies can, in principle, be calculated from a long equilibrium simulation where many binding-unbinding events occur, spontaneous protein-phospholipid membrane unbinding events are not observed in this study and have, to our knowledge, not been reported in the literature.

FVIII C1 exhibits structural flexibility and has multimodal orientations in the membrane-bound state.

While the lipid interaction of FVIII C2 resembles that of FVa C2(45), FVIII C1 is markedly different in two ways: the trajectories do not converge to a well-defined unimodal binding orientation (Fig. 4, *bottom*), and FVIII C1 in the bound state is characterized by shape modulations and disorder in the coiled main chain leading to spike 1 (Fig. 5, *bottom*). There are two major contributions to the retardation of the binding motif that allow the FVIII C1 domain to adopt a variety of moderately to highly tilted molecular orientations in the membrane-bound state: 1) the loss of local structural integrity (Fig. S3) and 2) the fact that S1 in FVIII C1 is four residues shorter than in FVIII C2, making it extrude significantly less (Fig. 1A). The domain tilt angle is in general steeper for FVIII C2 than for FVIII C1; the averages of the angles for 4 individual C1/HMMM trajectories range between 30 and 65 degrees (Fig. 4, *bottom*).

Physically, the structural change in C1 is facilitated by the breaking of the main chain H-bonds between Q2036 and K2072, which in the X-ray crystallographic structure clamp the surrounding loop region to the body of the domain (Fig. S3). Upon release, the main chain of the region centered around residue 2030 adheres to the membrane bilayer (Fig. 5, *bottom*) and interacts favorably in a non-specific manner. Such a structural mechanism or conformational change could explain, in part, why a complex multiphase binding mechanism is observed for FVIII.(46) The general trend of the calculated order parameters clearly show that FVIII C1 is more flexible than FVIII C2 (Fig. 5), in particular at regions that are in contact with the membrane (including the spikes, and the loop region leading to spike 1). Surprisingly, FVIII C1 also exhibits increased flexibility in a loop (residues 2115-2123), which is never observed to interact with the membrane. The fact that FVIII C1 is relatively disordered (for such a β-rich domain) could further, apart from a historical interest in FVIII C2, be involved in causing the scarcity of FVIII C1 domain structures available in the Protein Data Bank (PDB)(47). There is supporting experimental evidence that the protein segment in FVIII C1 involved in the above described structural change is indeed very flexible in the form of the crystallographic temperature factors (PDB entry codes 3cdz(17) and 4bdv(48)). From our simulations it is not possible to dismiss the

possibility that FVIII C2 undergoes a similar structural change under much larger time scales.

A search of PS-specific interactions in C1/HMMM reveals that (Fig. S4) on the corresponding key positions, R2163 (the counterpart to R2320 in C2; Fig. 3) is only observed to interact with the carboxylic acid group of the lipids in a single trajectory and R2220 in FVIII C2 has no counterpart in FVIII C1 (K2065 is the closest basic residue in alignment positioned two residues downstream and it interacts only non-specifically with either negatively charged functional group of the lipids).

Dynamical interpretation of hemophilic missense mutations and antibody epitopes related to membrane-binding.

Our results offer a dynamical interpretation of the available hemophilia A disease genotypes related to dysfunctional membrane association and binding of FVIIIa via its dual discoidin domains (described in the previous section), a novel approach made accessible by the use of the HMMM. Previous analyses have inferred on the functional-structural causation of hemophilic missense mutations based upon either homology models(49)(50) and/or X-ray crystallographic structures(51), both of which are static representations of protein structure in the absence of a membrane. Mutations of special interest are those positioned at the membrane interacting surface of the C2-like domains and include the following basic residues: R2052, R2090, R2159, R2163 (in C1), and R2320 (in C2). Caution must be taken when interpreting mutations. Potential pitfalls include not only usual statistical fallacies but also unknown interactions with other biochemical components. With these precautions we here adopt the palatable concept that a hemophilic disease missense mutation located in the membrane-interacting surface patch of FVIII C1 or FVIII C2 is potentially caused by improper membrane binding. As has been noted previously(51), it is remarkable that only few of the documented hemophilic missense mutations are positioned in the membrane-interacting patch (including the spikes) of the discoidin domains of FVIII (Fig. 5) and, specifically, only two mutants (V2223M, positioned in between spikes 2 and 3, and A2201P located in spike 1) was proposed to disrupt membrane binding. This could suggests that the major components necessary for sufficient membrane interaction for FVIIIa *in vivo* are non-specific contributions, such as

hydrophobic partitioning and electrostatics, consistent with the performed simulations from this study. In addition to this general observation, our results further implicate residues R2163 (in C1), R2220, and R2320 (in C2) as potentially critical in that they are the only basic residues capable of forming the structurally PS-specific interactions with the PS-type lipids. Of these three, R2163 and R2320 are known to give rise to hemophilic conditions upon mutation. (HAMSTeRs database; http://hadb.org.uk/)

Antibodies KM33 and ESH-4 are known to modulate FVIII cellular uptake and binding.(16)(52)(53)(54) KM33 abrogates the interaction between FVIII and phospholipid by binding to residue regions 2092-2093 (within S3 of FVIII C1) and 2158-2159 (within S4 of FVIII C1).(55) Analogously, ESH-4 binding epitopes have been narrowed down to residue regions 2192-2196 (N-terminal side of S1 of FVIII C2), 2210-2125 (containing S2 of FVIII C2), and 2313-2316 (S4 of FVIII C2).(56) A (third) FVIII antibody ESH-8 binds residue regions 2234-2238 (in between S2 and S3 of FVIII C2) and 2267-2270 (in between S3 and S4 of FVIII C2) and does not interfere with phospholipid binding. Our results show excellent agreement with the binding epitopes of antibodies KM33, ESH-4, and ESH-8 in that the two former have epitopes located entirely within the membrane-contacting surface patch of FVIII C1 or FVIII C2, while the latter binds solely to regions not in contact with the membrane (Fig. 5).

Proposed mechanism of membrane-interaction for the full-length FVIII and implications for a putative FVIIIa:FIXa tenase complex

Based on our results for the membrane-bound states of FVIII C1 and FVIII C2, putative membrane-bound configurations of the intrinsic tenase complex is proposed. Overlaying the putative model of the FVIIIa:FIXa teanse complex (Fig. 6A) with the converged final membrane-bound orientations of the C2/HMMM simulations (Fig. 6B, *top*) we find that, in all cases, preferential positioning of the C2 domain gives rise to tenase complex models without any clashes with the membrane. C1 domain of FVIII, however, is not in contact with the membrane, nor is the Gla-domain at (around) its ω-loop, as proposed for highly homologous Gla-domains of prothrombin(57) or FVIIa(58). On the contrary, when FVIII is overlaid the final FVIII C1 domain orientations from the C1/HMMM simulations (Fig. 6B, *bottom*), two of the resulting orientations causes significant clashes between FVIII C2 and

the membrane (trajectories #3 and #5). Trajectories #2 and #4 of the C1/HMMMs correspond to reasonable binding modes between the membrane and the tenase complex in which FVIII C1, FVIII C2 and FIXa Gla domains are anchored in the membrane (#2 looks intuitive where the entire complex is positioned normal to the membrane plane while #4 is very similar to highly tilted orientation suggested by FRET measurements(19)). Collectively, these considerations indicate that optimal interaction between FVIII C2 and the membrane can be achieved without major domain rearrangements in FVIII, whereas FVIII C1 membrane interactions might induce domain rearrangements in FVIII. Since it is known that the FVIII binding to membranes involves a complex mechanism(46), and considering the coiled nature of the four-residue FVIII C1-C2 inter-domain linker (residues 2170-2173), domain rearrangements are plausible. In any case, our simulations clearly indicate that FVIII C1 (not FVIII C2) plays an important role in modulating the orientation of the FVIII molecule to accommodate the interaction with FIXa in the membrane-bound state. The inter-domain flexibility of the full-length FVIIIa is unknown and, in particular, whether and how the membrane selects for or induces certain relative domain configurations. Furthermore, there appear to be non-trivial differences in PS-specificity for membrane-binding of FVIII and FVIIIa.(59) However, preliminary results of simulations of the crystal structure FVIII light chain bound to HMMM suggest that only minor relative domain rearrangements occur, supporting the validity of the presented rigid body putative tenase complex model (data not shown). These concepts are intriguing and subject for further studies.

Conclusion

The presented simulations show that peripheral tight binding and insertion of the domain extruding spikes occur spontaneously and converge to a domain orientation for human FVIII C2 that is nearly membrane-perpendicular. Human FVIII C1, however, undergoes a structural change and is orientationally promiscuous in that it can adopt one of several tilted orientations in the membrane-bound state. For both discoidin C2-like domains of FVIIIa, the mode of interaction with the phospholipid membrane is characterized by initial non-specific attractive electrostatic interactions and hydrophobic partitioning. The insertion of the membrane-anchoring domain is followed by the maturation of a tight PS-specific interaction where a lipid enters the PS pocket, which is observed more frequently for FVIII C2 than for FVIII C1. Deep embedding of large parts of either discoidin C2-like domain, as has been suggested for FVIII C2 based on cryo-electron microscopy, is not observed.

Acknowledgements

J.J.M. is supported by a grant from Forsknings- og Innovationsstyrelsen (Danish Agency for Science, Technology and Innovation) and the Novo Nordisk R&D Science, Talent, Attraction and Recruitment (STAR) programme.

References

1. Ohkubo, Y.Z., T. V Pogorelov, M.J. Arcario, G.A. Christensen, and E. Tajkhorshid. 2012. Accelerating membrane insertion of peripheral proteins with a novel membrane mimetic model. Biophys. J. 102: 2130–9.
2. MacFarlane, R. 1964. An enzyme cascade in the blood clotting mechanism, and its function as a biochemical amplifier. Nature. 202: 498–499.
3. Davie, E.W., and O.D. Ratnoff. 1964. Waterfall Sequence for Intrinsic Blood Clotting. Science. 145: 1310–1312.
4. Marder, V.J., W.C. Aird, J.S. Bennett, S. Schulman, and G.C. White. 2012. Hemostasis and Thrombosis: Basic Principles and Clinical Practice. Sixth edit. LWW.
5. Dieijen, G. Van, G. Tans, J. Rosing, and H.C. Hemker. 1981. The Role of Phospholipid and Factor VIIIa in the Activation of Bovine Factor X. J. Biol. Chem. 256: 3433–3442.
6. Lemmon, M.A. 2008. Membrane recognition by phospholipid-binding domains. Nat. Rev. Mol. Cell Biol. 9: 99–111.
7. Wakabayashi, H., A.E. Griffiths, and P.J. Fay. 2010. Factor VIII lacking the C2 domain retains cofactor activity in vitro. J. Biol. Chem. 285: 25176–84.
8. Sharon, N., and H. Lis. 2003. LECTINS. 2nd ed. Springer.
9. Mertens, K., and A.B. Meijer. 2012. Factors VIII and V swap fatty feet. Blood. 120: 1761–3.
10. Macedo-Ribeiro, S., W. Bode, R. Huber, M.A. Quinn-Allen, S.W. Kim, et al. 1999. Crystal structures of the membrane-binding C2 domain of human coagulation factor V. Nature. 402: 434–9.
11. Pratt, K.P., B.W. Shen, K. Takeshima, E.W. Davie, K. Fujikawa, et al. 1999. Structure of the C2 domain of human factor VIII at 1.5Å resolution. Nature. 402: 439–442.
12. Gilbert, G.E., R.J. Kaufman, A.A. Arena, H. Miao, and S.W. Pipe. 2002. Four hydrophobic amino acids of the factor VIII C2 domain are constituents of both the membrane-binding and von Willebrand factor-binding motifs. J. Biol. Chem. 277: 6374–81.

13. Gilbert, G.E., V.A. Novakovic, R.J. Kaufman, H. Miao, and S.W. Pipe. 2012. Conservative mutations in the C2 domains of factor VIII and factor V alter phospholipid binding and cofactor activity. Blood. 120: 1923–32.
14. Lü, J., S.W. Pipe, H. Miao, M. Jacquemin, and G.E. Gilbert. 2011. A membrane-interactive surface on the factor VIII C1 domain cooperates with the C2 domain for cofactor function. Blood. 117: 3181–9.
15. Hsu, T.-C., K.P. Pratt, and A.R. Thompson. 2008. The factor VIII C1 domain contributes to platelet binding. Blood. 111: 200–8.
16. Meems, H., A.B. Meijer, D.B. Cullinan, K. Mertens, and G.E. Gilbert. 2009. Factor VIII C1 domain residues Lys 2092 and Phe 2093 contribute to membrane binding and cofactor activity. Blood. 114: 3938–46.
17. Ngo, J.C.K., M. Huang, D.A. Roth, B.C. Furie, and B. Furie. 2008. Crystal structure of human factor VIII: implications for the formation of the factor IXa-factor VIIIa complex. Structure. 16: 597–606.
18. Liu, Z., L. Lin, C. Yuan, G.A.F. Nicolaes, L. Chen, et al. 2010. Trp2313-His2315 of factor VIII C2 domain is involved in membrane binding: structure of a complex between the C2 domain and an inhibitor of membrane binding. J. Biol. Chem. 285: 8824–9.
19. Wakabayashi, H., and P.J. Fay. 2013. Molecular orientation of Factor VIIIa on the phospholipid membrane surface determined by fluorescence resonance energy transfer. Biochem. J. 452: 293–301.
20. Parmenter, C.D.J., M.C. Cane, R. Zhang, and S. Stoilova-McPhie. 2008. Cryo-electron microscopy of coagulation Factor VIII bound to lipid nanotubes. Biochem. Biophys. Res. Commun. 366: 288–93.
21. Stoilova-McPhie, S., C.D.J. Parmenter, K. Segers, B.O. Villoutreix, and G.A.F. Nicolaes. 2008. Defining the structure of membrane-bound human blood coagulation factor Va. J. Thromb. Haemost. 6: 76–82.
22. Stoilova-McPhie, S., G.C. Lynch, S. Ludtke, and B.M. Pettitt. 2013. Domain organization of membrane-bound factor VIII. Biopolymers. 99: 448–459.
23. Li, H., A.D. Robertson, and J.H. Jensen. 2005. Very fast empirical prediction and rationalization of protein pKa values. Proteins. 61: 704–21.

24. Martínez, L., R. Andrade, E.G. Birgin, and J.M. Martínez. 2009. Packmol : A Package for Building Initial Configurations. J. Comput. Chem. 30: 2157–2164.
25. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14: 33–38.
26. Nagle, J.F., and S. Tristram-Nagle. 2000. Structure of lipid bilayers. Biochim. Biophys. Acta - Rev. Biomembr. 1469: 159–195.
27. Martyna, G.J., D.J. Tobias, and M.L. Klein. 1994. Constant pressure molecular dynamics algorithms. J. Chem. Phys. 101: 4177–4189.
28. Feller, S.E., Y.H. Zhang, R.W. Pastor, and B.R. Brooks. 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. 103: 4613–4621.
29. Allen, M.P., and D.J. Tildesley. 1989. Computer simulation of liquids. Reprint ed. Oxford University Press.
30. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98: 10089–10092.
31. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, et al. 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103: 8577–8592.
32. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al. 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26: 1781–1802.
33. Best, R.B., X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, et al. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. J. Chem. Theory Comput. 8: 3257–3273.
34. MacKerell, A.D., M. Feig, and C.L. Brooks. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J. Comput. Chem. 25: 1400–15.
35. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79: 926–935.

36. Ryckaert, J., G. Ciccotti, and H. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23: 327–341.
37. Cecchini, M., F. Rao, M. Seeber, and A. Caflisch. 2004. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. J. Chem. Phys. 121: 10748–56.
38. Lechtenberg, B.C., T.A. Murray-Rust, D.J. Johnson, T.E. Adams, S. Krishnaswamy, et al. 2013. Crystal structure of the prothrombinase complex from the venom of Pseudonaja textilis. Blood. 122: 2777–83.
39. Banner, D.W., A. D'Arcy, C. Chène, F.K. Winkler, A. Guha, et al. 1996. The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. Nature. 380: 41–46.
40. Brandstetter, H., M. Bauer, R. Huber, P. Lollar, and W. Bode. 1995. X-ray structure of clotting factor IXa: active site and module structure related to Xase activity and hemophilia B. Proc. Natl. Acad. Sci. U. S. A. 92: 9796–800.
41. Novakovic, V.A., D.B. Cullinan, H. Wakabayashi, P.J. Fay, J.D. Baleja, et al. 2011. Membrane-binding properties of the Factor VIII C2 domain. Biochem. J. 435: 187–96.
42. Lomize, M.A., A.L. Lomize, I.D. Pogozheva, and H.I. Mosberg. 2006. OPM: orientations of proteins in membranes database. Bioinformatics. 22: 623–5.
43. Shen, B.W., P.C. Spiegel, C.-H. Chang, J.-W. Huh, J.-S. Lee, et al. 2008. The tertiary structure and domain organization of coagulation factor VIII. Blood. 111: 1240–7.
44. Pantazatos, D., C.R. Gessner, V.L. Woods Jr, and G.E. Gilbert. 2014. Changes in the Factor VIII C2 Domain upon Membrane Binding Determined by Deuterium Exchange Mass Spectroscopy. Biochem. J. 461: 443–451.
45. Ohkubo, Y.Z., and E. Tajkhorshid. 2013. Exploring Membrane-Bound form of the C2 Domain by HMMM Model. Biophys. J. 104: 432a.
46. Bardelle, C., B. Furie, B.C. Furie, and G.E. Gilbert. 1993. Membrane binding kinetics of factor VIII indicate a complex binding process. J. Biol. Chem. 268: 8815–24.

47. Bernstein, F.C., T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, et al. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. 112: 535–542.
48. Svensson, L.A., L. Thim, O.H. Olsen, and E.M. Nicolaisen. 2013. Evaluation of the metal binding sites in a recombinant coagulation factor VIII identifies two sites with unique metal binding properties. Biol. Chem. 394: 761–5.
49. Pellequer, J.L., A.J. Gale, J.H. Griffin, and E.D. Getzoff. 1998. Homology models of the C domains of blood coagulation factors V and VIII: a proposed membrane binding mode for FV and FVIII C2 domains. Blood Cells. Mol. Dis. 24: 448–61.
50. Gale, A.J., J.L. Pellequer, E.D. Getzoff, and J.H. Griffin. 2000. Structural basis for hemophilia A caused by mutations in the C domains of blood coagulation factor VIII. Thromb. Haemost. 83: 78–85.
51. Liu, M.L., B.W. Shen, S. Nakaya, K.P. Pratt, K. Fujikawa, et al. 2000. Hemophilic factor VIII C1- and C2-domain missense mutations and their modeling to the 1.5-angstrom human C2-domain crystal structure. Blood. 96: 979–87.
52. Meems, H., M. van den Biggelaar, M. Rondaij, C. van der Zwaan, K. Mertens, et al. 2011. C1 domain residues Lys 2092 and Phe 2093 are of major importance for the endocytic uptake of coagulation factor VIII. Int. J. Biochem. Cell Biol. 43: 1114–21.
53. Wroblewska, A., S.D. van Haren, E. Herczenik, P. Kaijen, A. Ruminska, et al. 2012. Modification of an exposed loop in the C1 domain reduces immune responses to factor VIII in hemophilia A mice. Blood. 119: 5294–300.
54. Brinkman, H.-J.M., K. Mertens, and J.A. van Mourik. 2002. Phospholipid-Binding Domain of Factor VIII Is Involved in Endothelial Cell-Mediated Activation of Factor X by Factor IXa. Arterioscler. Thromb. Vasc. Biol. 22: 511–516.
55. Bloem, E., M. van den Biggelaar, A. Wroblewska, J. Voorberg, J.H. Faber, et al. 2013. Factor VIII C1 Domain Spikes 2092-2093 and 2158-2159 Comprise Regions That Modulate Cofactor Function and Cellular Uptake. J. Biol. Chem. 288: 29670–9.
56. Faber, J.H., R.S. Appa, H. Stennicke, and M. Kjalke. 2013. Contribution of C2 domain epitopes in FVIII cellular uptake assessed by epitope mapping of the anti-C2 antibodies ESH4 and ESH8 by hydrogen-deuterium exchange mass spectrometry.

In: FASEB Summer Research Conference on Proteases in Hemostasis & Vascular Biology. June 2–7, Nassau, Bahamas.

57. Huang, M., A.C. Rigby, X. Morelli, M.A. Grant, G. Huang, et al. 2003. Structural basis of membrane binding by Gla domains of vitamin K-dependent proteins. Nat. Struct. Biol. 10: 751–6.

58. Ohkubo, Y.Z., and E. Tajkhorshid. 2008. Distinct structural and adhesive roles of Ca2+ in membrane binding of blood coagulation factors. Structure. 16: 72–81.

59. Engelke, H., S. Lippok, I. Dorn, R.R. Netz, and J.O. Rädler. 2011. FVIII binding to PS membranes differs in the activated and non-activated form and can be shielded by annexin A5. J. Phys. Chem. B. 115: 12963–70.

Figure- and table legends

FIGURE 1: (**A**) Structural alignment of the domains C1 (wire representation, pink) and C2 (wire representation, cyan) from FVIII based on PDB entry code 3cdz(17). The spikes S1-S4 are highlighted in red for FVIII C1 and in blue for FVIII C2. (**B**) Initial domain orientation of the C2-like domain and the membrane patch in the simulation setup. Depicted is FVIII C2 (in wire representation) and the HMMM patch (half-transparent). The vectors (yellow) show the principal axes of inertia of the FVIII C2 domain. The domain tilt angle, θ, is defined as the angle between the third principal axis of inertia and the membrane normal, *z*. (**C**) Converged FVIII C2 (trajectory #2) in the membrane-bound state with inserted spikes. The vectors (yellow) show the principal axes of inertia of the FVIII C2 domain.

FIGURE 2: (*Left*) Insertion depth of spikes 1-3 of the FVIII C2 domain and interaction energy with the HMMM as function of simulation time for the five trajectories, #1-#5. Each spike depth is represented by a single α-carbon atom. The chosen residues are G2044 (Spike 1), S2058 (Spike 2), and F2093 (Spike 3). Gray-shaded is the layer below the average *z*-position of the phosphorus atoms of st-lipids, corresponding to the membrane region. The reference point (i.e., z=0) is set at the membrane center, which is defined in each trajectory frame as the center of mass of DCLE molecules. Calculated non-bonded interaction energy is divided into van der Waals (vdW, in green) and electrostatic (Elec, in blue) components. (*Right*) Insertion depth of spikes 1-3 of the FVIII C1 domain and interaction energy with the HMMM as function of simulation time for the four binding trajectories, #2-#5, (black) and the non-binding trajectory, #1 (red). Each spike depth is represented by a single α-carbon atom. The chosen residues are M2199 (Spike 1), R2215 (Spike 2), and L2252 (Spike 3). Gray-shaded is the layer below the average *z*-position of the phosphorus atoms of st-lipids, corresponding to the membrane region. The reference point (i.e., z=0) is set at the membrane center, which is defined in each trajectory frame as the center of mass of DCLE molecules. Calculated non-bonded interaction energy is divided into van der Waals (vdW, in green) and electrostatic (Elec, in blue) components.

FIGURE 3: (*Left*) Position of the R2220 and R2320 of FVIII C2 (cyan wire representation, spikes highlighted in blue). (*Right*) A cross section view of the FVIII C2

(cyan wire representation, spikes highlighted in blue, and black half-transparent surface) with a tightly bound st-lipid interacting with the R2320 of the characteristic soft 'funnel' PS pocket. The view point is 90 degree rotated towards right from the left panel. Snapshot was taken from trajectory #5 at t = 21 ns.

FIGURE 4: Distribution of the domain tilt angle, θ, of the FVIII C2 (top) and FVIII C1 (bottom) for the pooled subset of membrane-binding trajectories (#1-#5 for FVIII C2, and #2-#5 for FVIII C1). Domain tilt angles corresponding to the final converged configurations are indicated by arrows.

FIGURE 5: (*Top*) Calculated amide (C-N vector) S^2 order parameters (black circles) and membrane contact frequencies (brown shaded bars) for FVIII C2 in the membrane-bound state. The positions of the spikes (S1-S4) are indicated with blue colored bars above the panel. Hemophilic missense mutations are marked with crosses on the baseline. (*Bottom*) Calculated amide (C-N vector) S^2 order parameters (black circles) and membrane contact frequencies (brown shaded bars) for FVIII C1 in the membrane-bound state. The positions of the spikes (S1-S4) are indicated with red colored bars above the panel. Hemophilic missense mutations are marked with crosses on the baseline.

FIGURE 6: (**A**) Our constructed model of the FVIIIa:FIXa tenase complex. B-domain deleted FVIIIa (domains A1-A2-A3-C1-C2) is shown in wire representation (A1-A2 in blue, A3-C1-C2 in tan). FIXa is shown as surface (HC: Heavy Chain in gray, LC: Light Chain in green). (**B**) The tenase complex model structurally aligned and overlaid the final converged membrane-bound individual C2-like domains (*top*: C2/HMMM trajectories #1-#5 from left to right, *bottom*: C1/HMMM trajectories #2-#5 from left to right).

TABLE 1: Detailed overview and component count of the simulated systems.

Tables

System	Lipid type	Area / lipid [Å^2]	No. of trajectories	Spontaneous binding in	Simulation time per trajectory [ns]	No. of water molecules	No. of atoms
C1/HMMM	PS	88.5	5	All except #1	50-55	~7,800	~37,000
C2/HMMM	PS	85.0	5	All	~37	~7,700	~37,000
C2/Soln	-	-	1	-	50	~4,600	~16,400

TABLE 1: System configurations

Figures

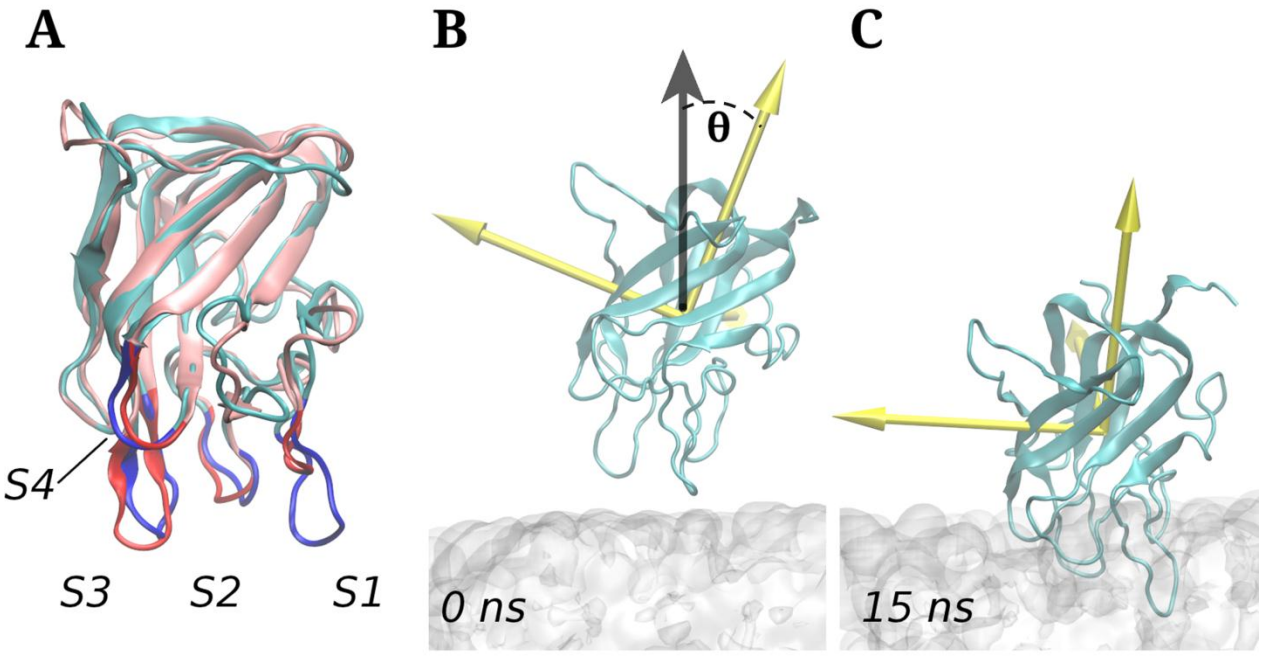


FIGURE 1

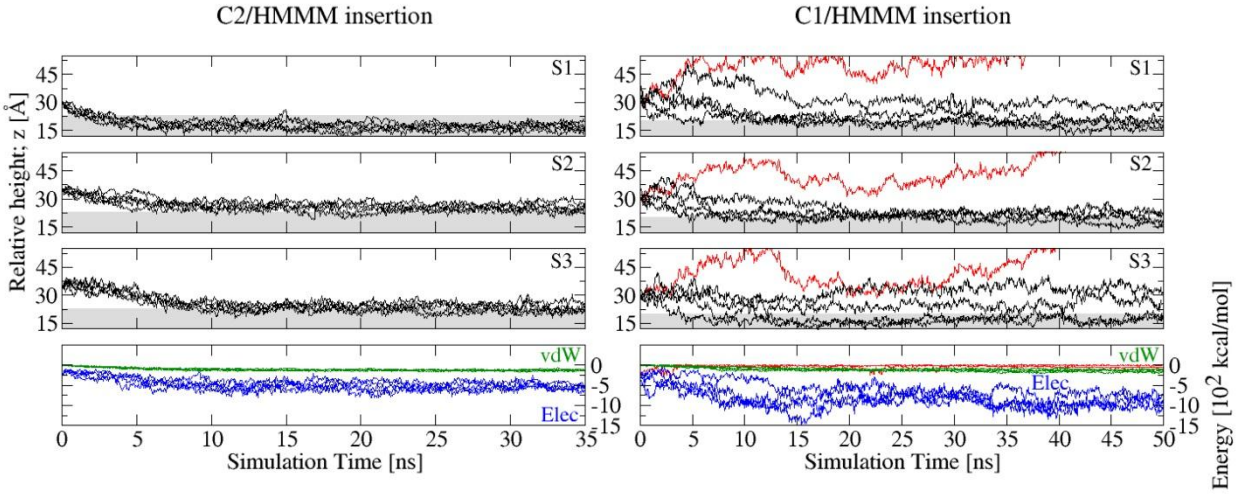

C2/HMMM insertion
C1/HMMM insertion
Relative height; z [Å]
S1
S2
S3
vdW
Elec
Simulation Time [ns]
Energy [10^-2 kcal/mol]

FIGURE 2

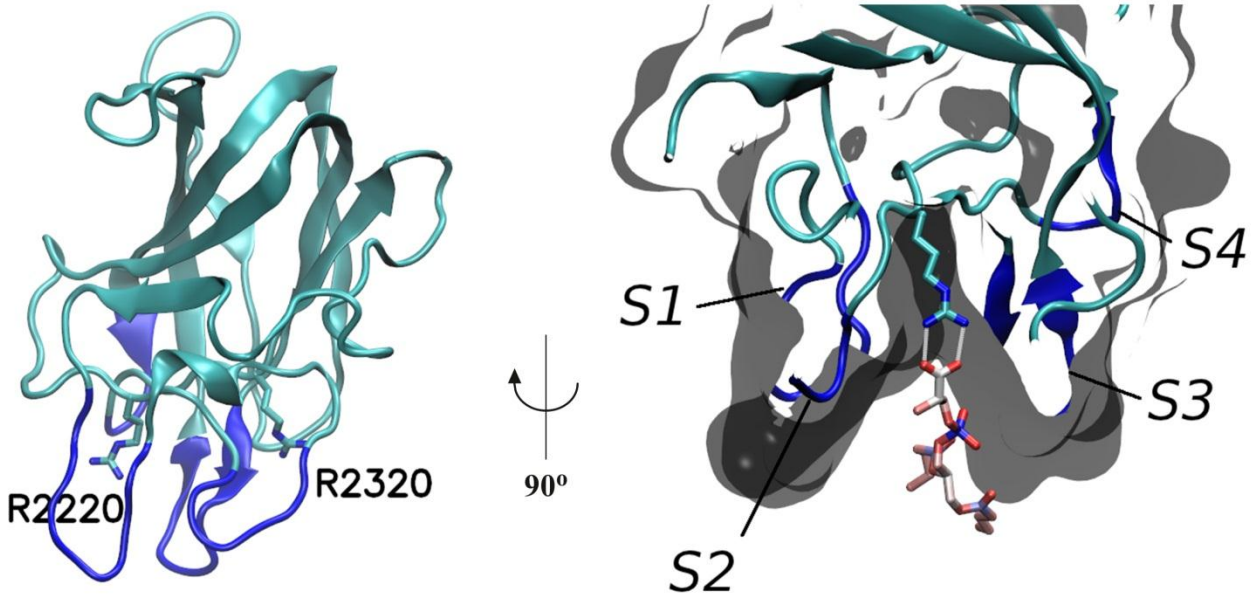


FIGURE 3

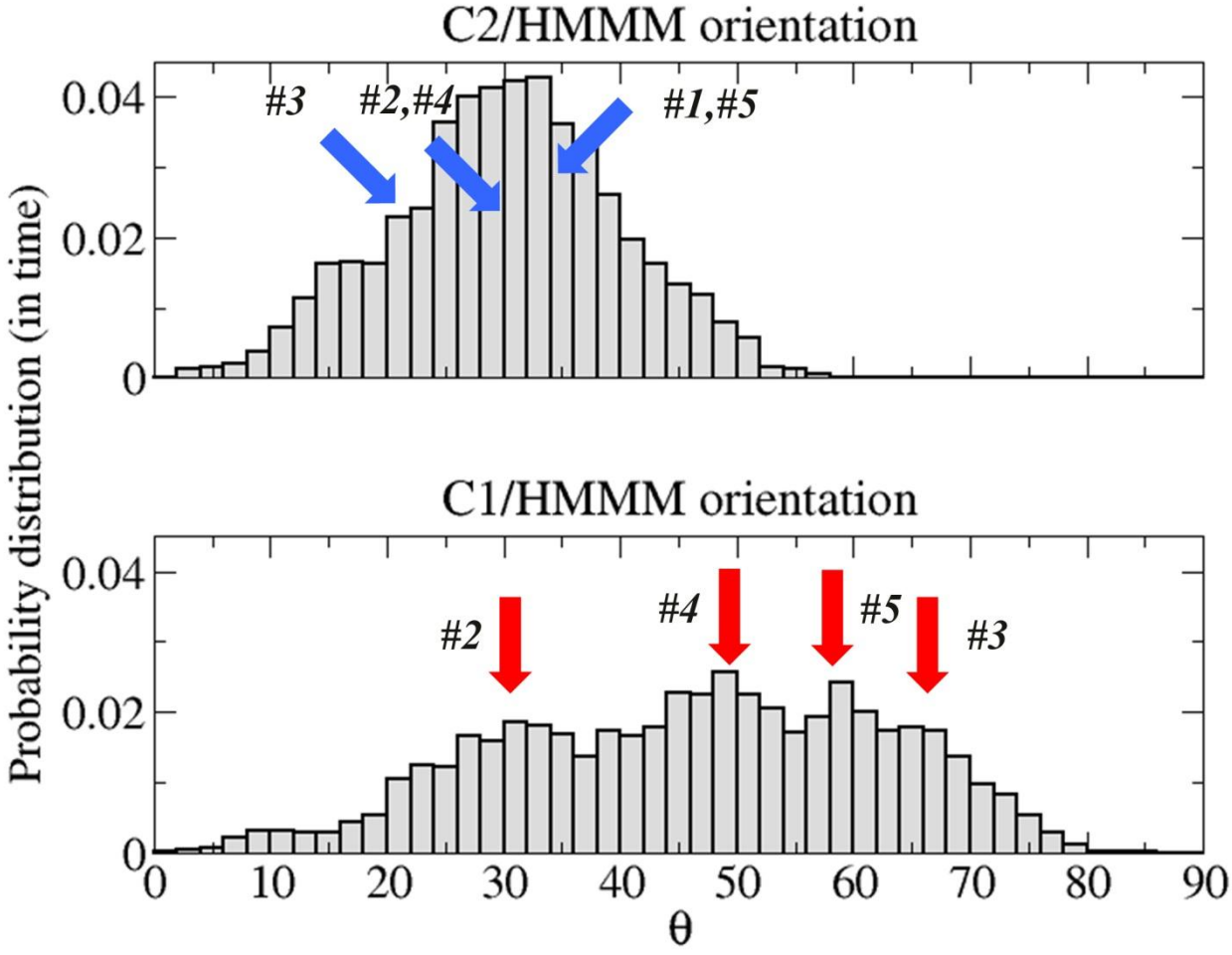

C2/HMMM orientation
#3
#2,#4
#1,#5
0.04
0.02
0
C1/HMMM orientation
#2
#4
#5
#3
0.04
0.02
0
0
10
20
30
40
50
60
70
80
90
θ
Probability distribution (in time)

FIGURE 4

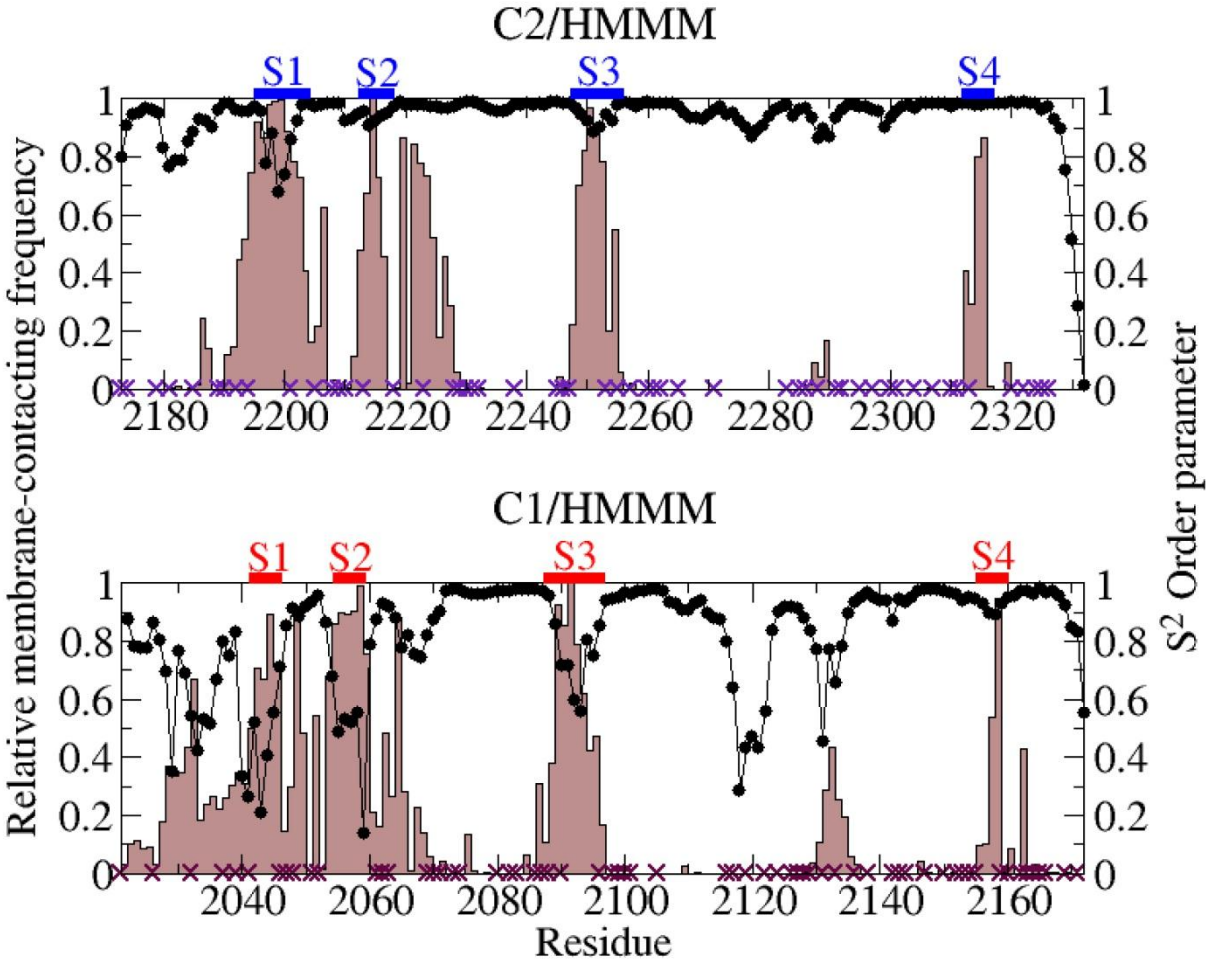

C2/HMMM
S1
S2
S3
S4
C1/HMMM
S1
S2
S3
S4
Relative membrane-contacting frequency
S² Order parameter
Residue
2180
2200
2220
2240
2260
2280
2300
2320
2040
2060
2080
2100
2120
2140
2160
0
0.2
0.4
0.6
0.8
1

FIGURE 5

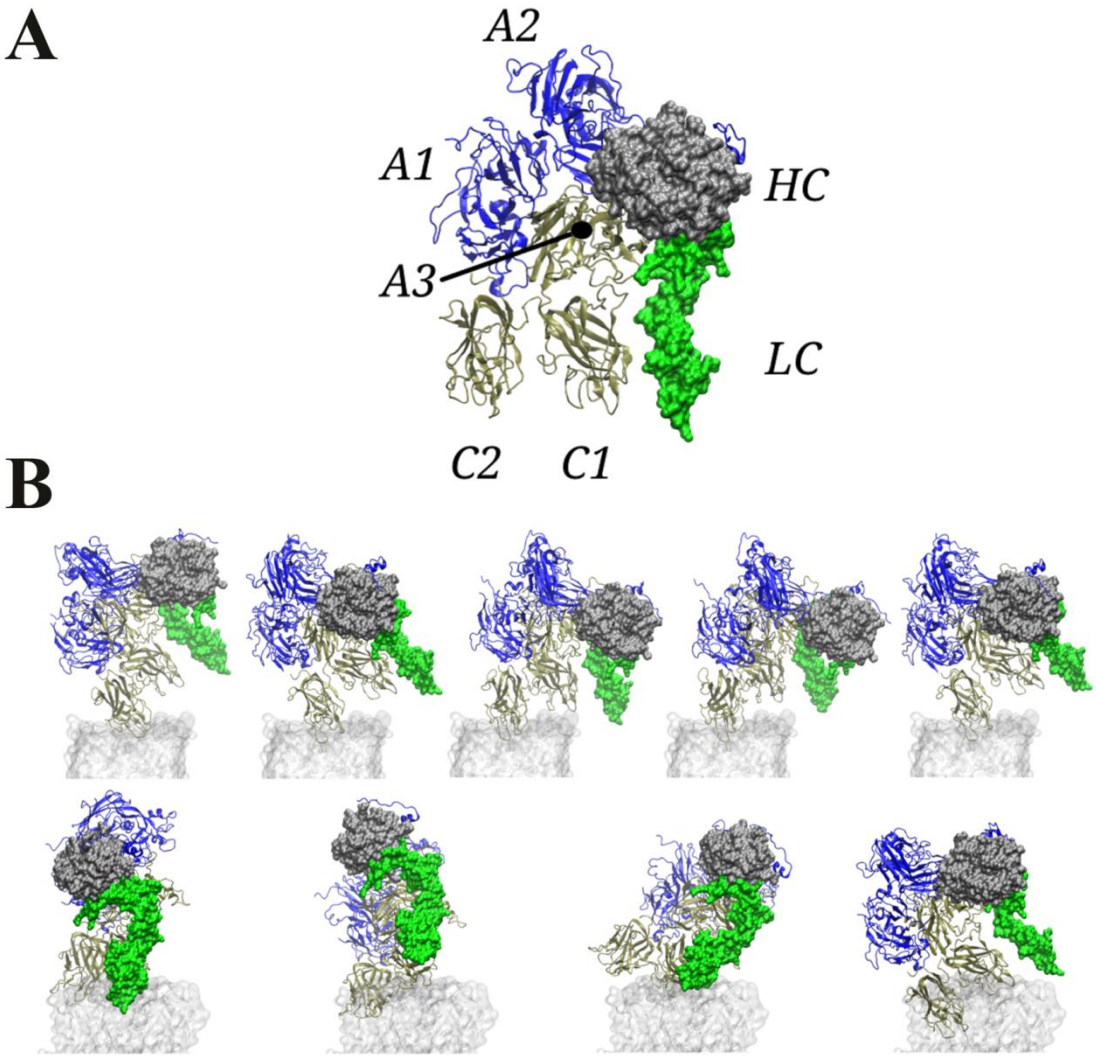


FIGURE 6

Supporting Material

Membrane-interaction of the factor VIIIa discoidin domains: implications for the tenase complex

Jesper J. Madsen[1,2], Y. Zenmei Ohkubo[3], Emad Tajkhorshid[3,4], Günther H. Peters[2], Johan H. Faber[5], Ole H. Olsen[1]

[1] Haemophilia Biochemistry, and [5] Protein Characterisation, Novo Nordisk A/S, DK-2760 Måløv, Denmark.

[2] Department of Chemistry, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

[3] Beckman Institute for Advanced Science and Technology, and [4] Center for Biophysics and Computational Biology, School of Chemical Sciences, Departments of Biochemistry and Chemistry, College of Medicine, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

Correspondence: Ole H. Olsen, Haemophilia Biochemistry, Novo Nordisk A/S, Novo Nordisk Park, F5.1.29, DK-2760 Måløv, Denmark.
Tel.: +45 3075 4511
E-mail: oho@novonordisk.com

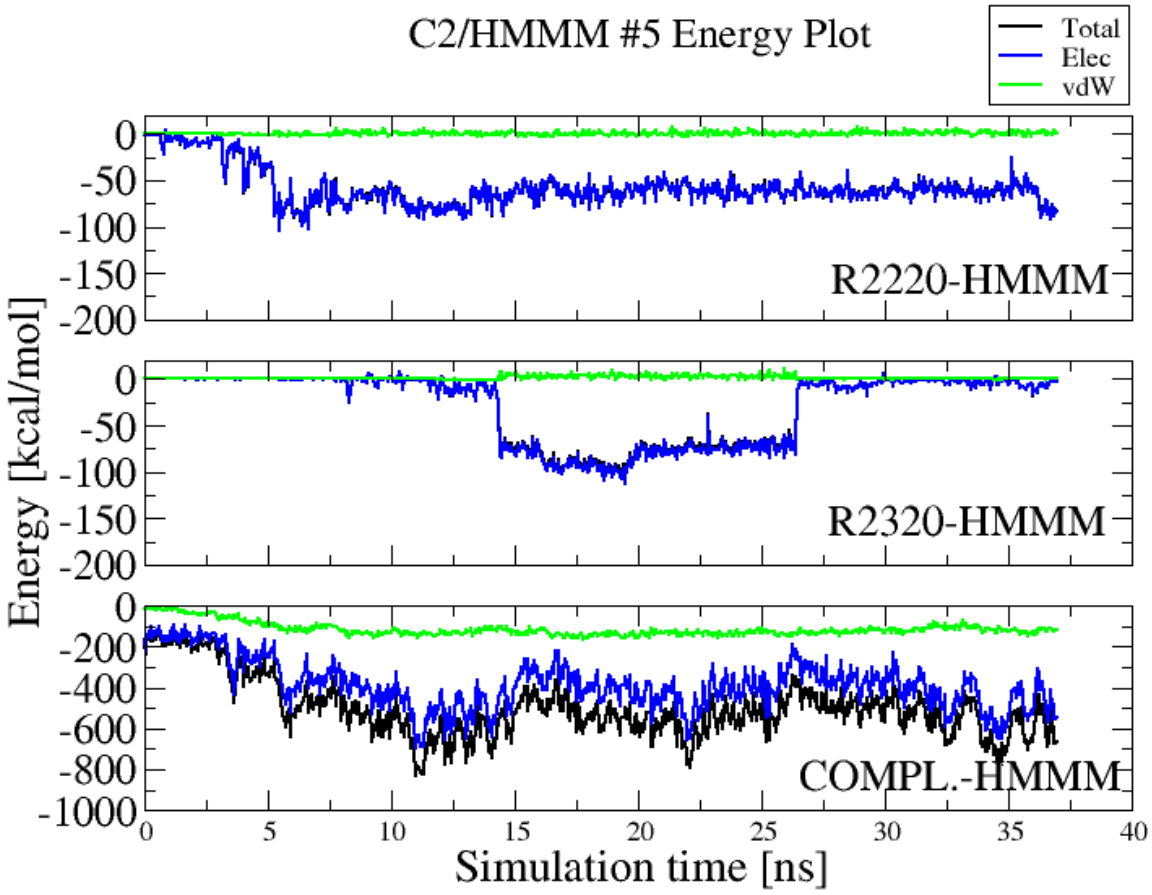


Figure S1: Non-bonded interactions energies between the HMMM membrane and specific portions of C2 domain: R2220 (Top panel), R2320 (Center), or the complementary selection (i.e., C2 domain excluding R2220 and R2320; Bottom) for a FVIII C2 on HMMM trajectory (C2/HMMM #5). The energies are decomposed into electrostatic (blue) and van der Waals (green) terms; the total (shown in black) is the sum of the two terms.

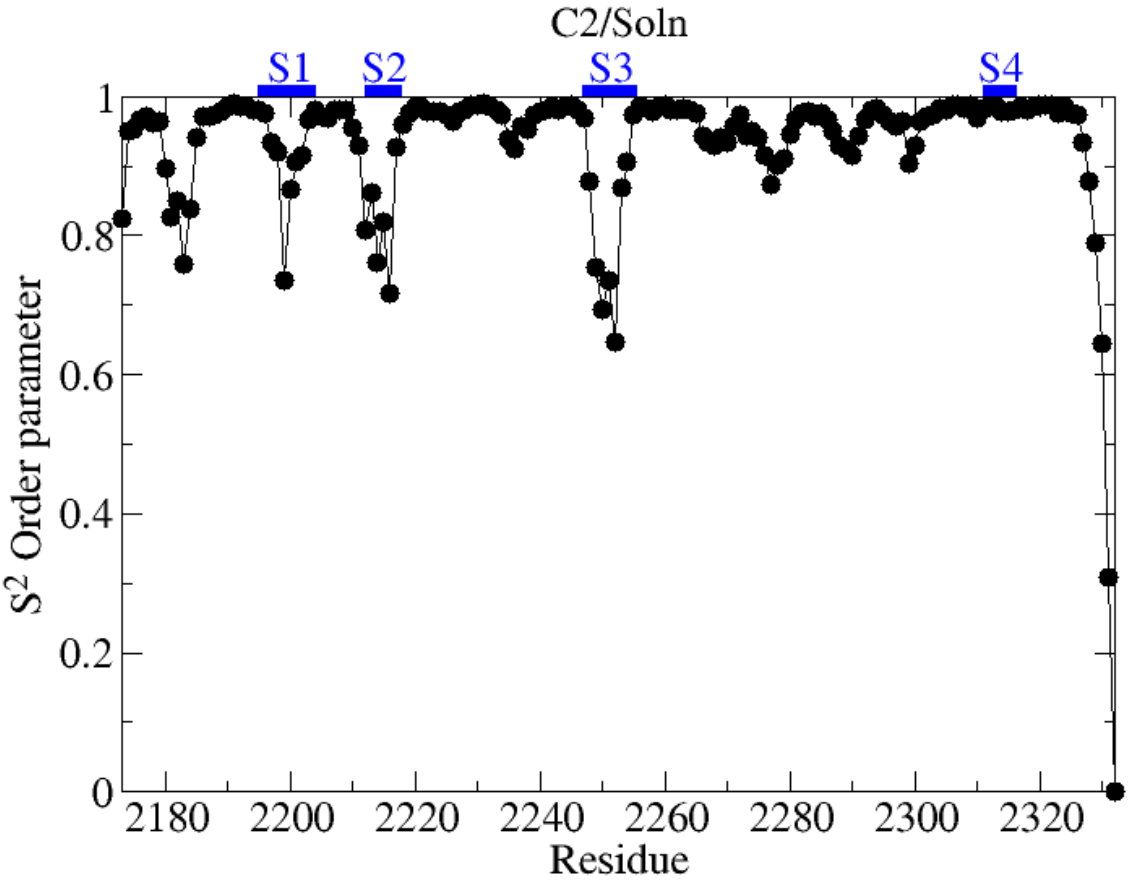


Figure S2: Calculated amide (C-N vector) S^2 order parameters for human FVIII C2 in bulk water (C2/Soln trajectory).

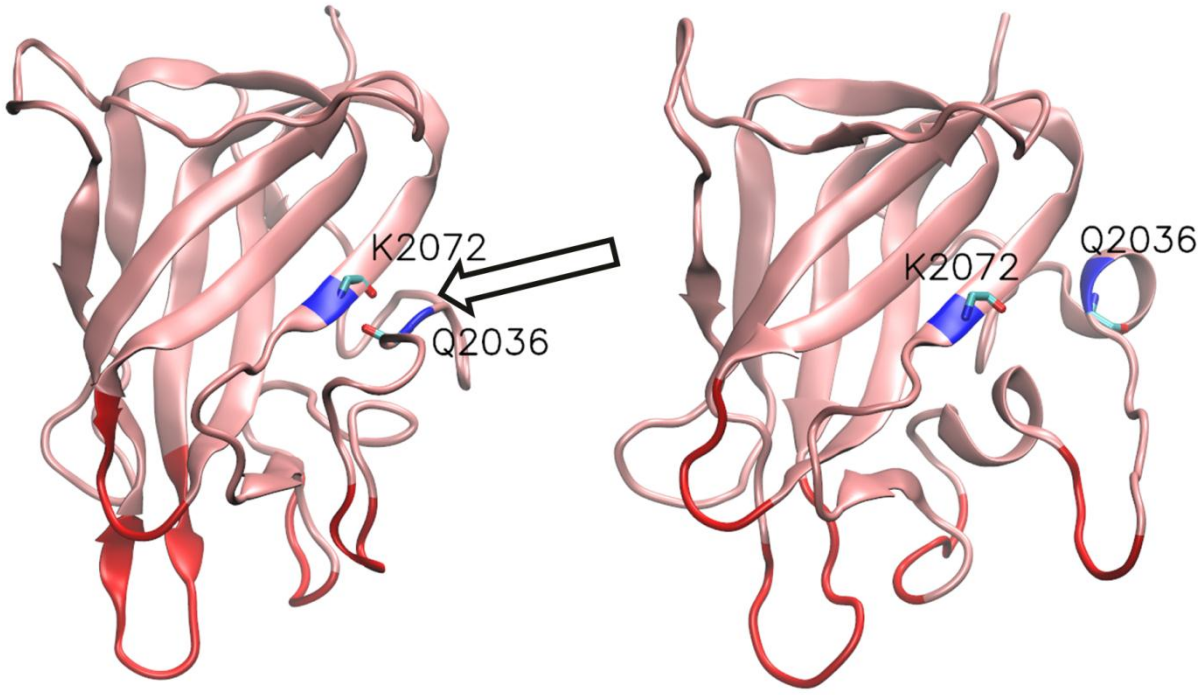


Figure S3: Structural change in C1 upon the release of the main chain leading to spike 1 from the body of the β-barrel. The C1 domain is shown in wire representation (pink) with the spikes 1-4 highlighted (red). Positions of residues Q2036 and K2072 which restrain the loopy segment in the X-ray crystallographic structure are indicated in blue and backbone atoms are shown as sticks. (*Left*) X-ray crystallographic reference structure. (*Right*) Snap-shot from C1/HMMM trajectory #2 at t = ~32 ns.

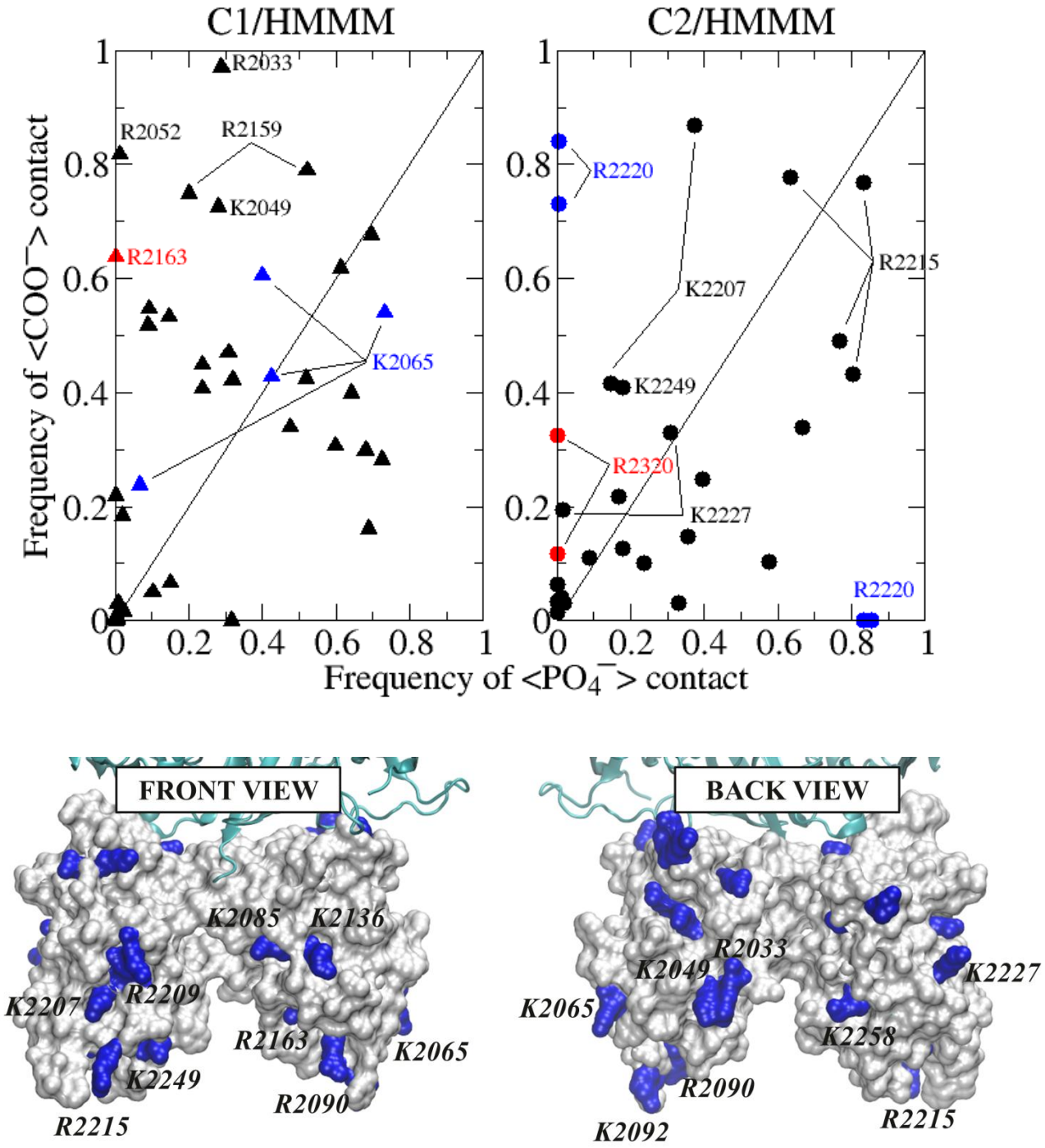


Figure S4: (*Top*) Observed frequency of close interaction (cutoff = 3 Å) between basic residues and negatively charged functional groups (PO_4^- along the x-axis; COO^- along the y-axis) of the PS head groups. Each point in the plot corresponds to the interaction frequency between single residue averaged over a single membrane-binding trajectory of

the individual C2-like domains (*Top Left*: FVIII C1; *Top Right*: FVIII C2.) (*Bottom*) Front-side and back-side view of FVIII C1 and C2 shown as surface (PDB entry code 3cdz) with basic residues (K or R) highlighted (blue). Front view corresponds to the orientation of Fig. 6A with FVIII C1 positioned to the right and FVIII C2 to the left.

4.2 Clearance of FVIIIa by LRP1

4.2.1 Prelude

The final manuscript of this dissertation, manuscript 5, describes a detailed biochemical and biophysical characterization of the binding between LRP1 of the LDL receptor family and (one of) its natural ligands, coagulation factor VIII. LRP1 is known to clear the FVIII molecule from circulation *in vivo*(Bovenschen et al., 2003) and lysines in the ligand well-established as being critical for the interaction with the receptor.(Dolmer, Campos, & Gettins, 2013; Fisher, Beglova, & Blacklow, 2006) Other physiological clearance receptors of FVIII include proteoglycans and carbohydrate receptors.(Lenting, VAN Schooten, & Denis, 2007) The typical half-life of FVIII in the bloodstream is ~10 hours, although numbers vary significantly among patients with haemoplilia A.(Dijk et al., 2005)

The LRP receptor contains cytoplasmic tail, a transmembrane-spanning segment, auxiliary domains, and the ligand-binding complement-type cysteine-rich repeasts (CRs). The principal binding cluster (cluster II) has 8 of these CR repeats numbered 3 through 10 (of 31 in the full-length receptor). This modular composition where many repeats can interact with different regions of the ligand enables LRP to be a promiscuous receptor and achieve affinity through avidity.

Based on selected anchor-point lysine residues identified in the mutagenesis study, an illustrative model of the canonical interaction between a LRP1 cluster II CR-triplet(Dolmer & Gettins, 2006)(Meijer et al., 2007) and an extended surface patch also covering regions with reduced deuterium-incorporations in hydrogen-deuterium exchange mass spectrometry experiments on FVIII was constructed.

4.2.2 Manuscript 5: van den Biggelaar M, *et al.*, to be submitted

Low-Density Lipoprotein Receptor-Related Protein 1 Interacts with its Ligand Blood Coagulation Factor VIII via an Extended Surface Comprising Multiple Lysine Residues

M. van den Biggelaar[1], J.H. Faber[2], J.J. Madsen[2], M. Zuurveld[1], C. van der Zwaan[1], O.H. Olsen[2], H.R. Stennicke[2], K. Mertens[1,3], A.B. Meijer[1,3]

[1]From the Department of Plasma Proteins, Sanquin Research, Amsterdam, The Netherlands

[2]From the Biopharmaceutical Research Unit, Novo Nordisk A/S DK-2760 Måløv, Denmark

[3]From the Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

*Running title: *Lysine residues mediate LRP1-FVIII binding*

To whom correspondence should be addressed: Dr. Maartje van den Biggelaar, Department of Plasma Proteins, Sanquin Research, Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands, Tel.: +31-20-5123123; Fax: +31-20-5123310; E-mail: m.vandenbiggelaar@sanquin.nl

Keywords: Coagulation factors, Receptors, Mass spectrometry, Surface plasmon resonance, Protein-protein interactions, Factor VIII, Low-density lipoprotein receptor-related protein, Hydrogen-deuterium exchange

---

Background: Lysine residues mediate the interaction between ligands and members of the LDL receptor family.

Results: Regions on FVIII and specific lysine residues that contribute to the interaction of FVIII and LRP1 are identified.

Conclusion: LRP1 interacts with FVIII via an extended surface comprising multiple lysine residues. **Significance:** Our study sheds light on ligand interaction with the LDL receptor family.

ABSTRACT

Lysine residues are implicated in driving the ligand binding to the LDL receptor family. However, it has remained unclear how specificity is regulated. Using coagulation factor VIII as a model ligand, we now study the contribution of individual lysine residues in the interaction with the largest member of the LDL receptor family, low-density lipoprotein receptor-related protein (LRP1). Using hydrogen-deuterium exchange mass spectrometry (HDX-MS) and surface plasmon resonance (SPR) interaction analysis on a library of lysine replacement variants as two independent approaches, we demonstrate that the interaction between Factor (F)VIII and LRP1 occurs over an extended surface containing multiple lysine residues. None of the individual lysine residues account completely for LRP1 binding, suggesting a cooperative binding model. Together with structural docking studies, our data suggest that LRP1 interacts with FVIII via an extended surface of multiple lysine residues that starts at the bottom of the C1 domain and winds around the FVIII molecule.

The low-density lipoprotein receptor family is an ancient family of transmembrane receptors regulating diverse biological processes ranging from lipoprotein metabolism to blood coagulation(1-3). Members of the LDL receptor family are able to interact with an unusual number of ligands; e.g. more than 30 ligands have been reported for the largest member LRP1, including lipoproteins, viruses, bacteria toxins, proteinase-inhibitor complexes and coagulation proteins(4,5). Given the importance of the LDL receptor family in human physiology, much research has been focused on the questions how the various members of the LDL receptor family can interact with such a large variety of structurally unrelated ligands and which molecular mechanisms regulate specificity of the interaction.

The LDL receptor family is characterized by the presence of a varying number of distinct domains; the YWTD β-propeller and EGF domains, a transmembrane region, an intracellular cytoplasmic tail and ligand-binding domains or complement-type repeats (CR domains). Ligand binding is mediated by clusters of structured CR domains that consist of approximately 40 amino acids folding into a compact structure. Each CR domain contains three disulfide bonds, a highly conserved octahedral calcium binding cage and a short β-hairpin near the N-terminal end(6). The calcium

cage is required both for structural integrity of the CR domain as well as for direct ligand binding and is formed by 6 oxygen atoms derived from four conserved acidic residues and two backbone carbonyl groups(6-9). The CR domains are connected by flexible linker sequences, thereby enabling the receptor family to bind to a wide range of proteins.

Most of our knowledge on ligand-receptor interaction is based on the intracellular chaperone receptor-associated protein (RAP) which binds with high affinity to members of the LDL receptor family(10,11). Based on a co-crystallization study of two CR domains from the LDL receptor with the third domain of RAP (RAP-D3), a general mode for ligand recognition by lipoprotein receptors has been proposed(12). In this 'acidic necklace' model, each CR domain encircles an ε-amino group of a lysine residue in a tripartite salt bridge (K256 and K270 in RAP-D3) via the three remaining oxygen atoms from the acidic residues forming the octahedral calcium cage(12). This model was supported by a previous study in which K256 and K270 were identified as being critical for the interaction of RAP-D3 to complete LRP1 using random mutagenesis(13). Further support for the lysine-binding model was provided by the NMR structure of the complex of complement type repeat 5 and 6 of LRP1 and the first domain of RAP (RAP-D1)(14), although the models differ in terms of the additional contribution of hydrophobic interactions. Recently, we further experimentally demonstrated that positively charged arginines cannot substitute for the two dominant lysine residues (K256 and K270) in the interaction with the RAP-D3 domain and LRP1(15). In agreement with these data, arginine residues do also not take over lysine residues in the high-affinity interactions of RAP-D1 (K143 and K146) with the LDL receptor(16). In a recent elegant study, Dolmer *et al*. established that lysine residues are the sole contributors to binding of RAP-D3 to a LRP1 fragment containing 2 CR domains and that pairs of lysine residues ensure high-affinity interaction in an additive rather than synergistic manner(17).

RAP prevents premature ligand-binding during biosynthesis of the LDL receptor family members by efficiently competing with other ligands(18-20). It therefore seems well possible that the affinity and molecular mechanism of interaction have been optimized during evolution. While the acidic necklace model is consistent with previous observations that lysine residues are implicated in the binding of ligands other than RAP to the LDL receptor family(21-28), it has remained unclear how large ligands containing multiple lysine residues interact with the LDL receptor family members. This is particularly relevant since experiments using single CR domains and simple (modified) amino acids have shown that also arginine and protonated histidine, especially pairs of proximal charges, can interact well with CR domains(29). Taken together, it has remained unclear whether lysine residues make a dominant contribution to the interaction with the LDL receptor family,

also for low-affinity ligands. Another important question is how specificity of the interaction is attained. It has been suggested that high-affinity interaction requires the engagement of at least two separate lysine residues located at an appropriate distance from each other(12,30). In addition, it has been proposed that there are modest requirements for being a coordinating lysine residue and that specificity may be regulated by a second, immediately adjacent positive charge(17,29). We therefore set out to study in detail the interaction of LRP1 with its low-affinity ligand FVIII using two independent approaches with particular emphasis on lysine residues. Using hydrogen-deuterium mass spectrometry we studied the interaction of B-domain deleted FVIII with LRP1 cluster II. Subsequently, we used a lysine mutagenesis approach to study the role of lysine residues in the FVIII light chain in the interaction with full length LRP1 and LRP1 cluster II. These combined studies revealed important novel aspects of the interaction between FVIII and LRP1.

EXPERIMENTAL PROCEDURES

Plasmid mutagenesis -FVIII light chain in pcDNA3.1 was created from wild type FVIII-GFP in pcDNA3.1(31) by fusing the signal peptide to the acidic *a3* domain (glutamic acid at position 1649) by Quick Change mutagenesis™ (Stratagene, La Jolla, CA, USA) using forward primer 5'-TTGCGATTCTGCTTTAGTGAAATAACTCGTACTACT-3' and reverse primer 5'-AGTAGTACGAGTTATTTCACTAAAGCAGAATCGCAA-3'. B domain deleted (del 746-1639) factor VIII (FVIIIdB) in pcDNA3.1 has been described before(32). Point mutations in the FVIII light chain and FVIIIdB were introduced by Quick Change mutagenesis™ (Stratagene, La Jolla, CA, USA) using appropriate primers. The coding regions of all constructs were verified by sequence analysis. Sequence reactions were performed with BigDye Terminator Sequencing kit (Applied Biosystems, Foster City, CA, USA).

Expression of recombinant proteins - Recombinant factor VIII (FVIII) Turoctocog alfa (previously named N8) was produced in Chinese hamster ovary (CHO) cells as previously described(33). The molecule consists of a heavy chain of 88 kDa including a 21 amino acid residue truncated B-domain and a light chain of 79 kDa, it contains four N-glycosylation sites of which two are complex bi-antennary glycans and two are high mannose structures. FVIII light chain variants were expressed under serum-free conditions using 293-F cells (Invitrogen, Carlsbad, CA, USA #R790-07) and Freestyle 293 Expression Medium (Invitrogen). 293Freestyle cells were transfected using 293fectin (Invitrogen) according to the manufacturer's instructions for small scale expression and using linear polyethylenimine (PEI) Mw~ 25000 (Polysciences Inc, PA, USA) for large scale expression. Cells were harvested 4-5 days after transfection. Cells were centrifuged for 8 min at 4600 × g and the supernatant was collected. The

cell pellet was subsequently washed with 0.55 M NaCl in Freestyle 293 Expression Medium and centrifuged for 8 min at 4600 × g. The conditioned medium was combined with the 0.55 M NaCl wash and filtered using a 0.22 µm disposable filter unit. FVIII light chain variants were purified using CLB-CAg117 coupled to CNBr-Sepharose 4B as an affinity matrix followed by concentration using Q-sepharose(31) (Amersham Pharmacia Biotech, Roosendaal, The Netherlands). HEK293 cell-lines stably expressing FVIIIdB variants were produced as described(34) and grown in DMEM-F12 medium supplemented with 10% FCS. Recombinant FVIIIdB variants were purified and analyzed as described using CLB VK34 IgG1 coupled to CNBr-Sepharose 4B as an affinity matrix followed by concentration using Q-sepharose(21). LRP1 cluster II was expressed in Baby Hamster Kidney (BHK) cells and purified as described(35) using glutathione S-transferase fused receptor-associated protein (RAP) coupled to CNBr-Sepharose 4B as an affinity matrix. Subsequently, purified LRP1 cluster II was concentrated using Q-sepharose.

HDX-MS experiments -189 µM FVIII and 256 µM LRP1 cluster II were dialyzed to a buffer comprising 20 mM imidazole pH 7.3, 10 mM $CaCl_2$, 500 mM NaCl and complex formation was allowed for 30 minutes at room temperature. In-exchange of deuterium was initiated by a 10-fold dilution into either 20 mM imidazole pH 7.3, 10 mM $CaCl_2$, 111 mM NaCl in 100% H_2O for the undeuterated experiments, or 20 mM imidazole pD 7.3, 10 mM $CaCl_2$, 111 mM NaCl in 100% D_2O for the deuterated experiments. All sample deuterium labeling, quenching, injection to MS and timing of samples were handled by an automated set-up (Leap Technologies Inc.) as described(36) with minor adjustments. Minor modifications included exchange times ranging from 10 seconds to 256 minutes, the use of a 7 min linear acetonitrile gradient of 8−45% containing 0.1% formic acid and the use of a SYNAPT G2 HDMS mass spectrometer (Waters). Four undeuterated and two complete HDX experiments were performed for FVIII and FVIII in complex with LRP1 cluster II within two consecutive days. Blank injections were performed between each sample injection to confirm the absence of peptide carryover from previous runs. Peptic peptides were identified in separate experiments using MS^E methods. MS^E data were processed using ProteinLynx Global Server software (Waters version 2.4). Searches of MS^E data were conducted as non-specific digest with N-terminal acetylation of FVIII as modification. The sequence coverage map of FVIII was plotted using the online tool MSTools(37). The relative deuterium incorporation levels for each peptic peptide were automatically calculated using DynamX software (Waters) as described(36). The data were expressed in either mass units (Da) or relative fractional exchange, which was calculated by dividing the deuterium level (in Da) by the total number of backbone amide hydrogens that could have become deuterated (equal to the number of

amino acids, minus proline residues minus 1 for the N-terminal amide). The experimental uncertainty of measuring a deuterium level was found to be ±0.15 Da. Using this value of experimental uncertainty and a 98% confidence interval, a significant difference value of 0.5 Da was calculated for individual time points and a value of 1.5 kDa for summed HDX differences across all time points.

Surface Plasmon Resonance analysis- Association and dissociation of LRP1 cluster II to FVIII light chain and FVIIIdB variants was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). The anti-C2 antibody CLB-EL14 IgG4(38) (26 fmol/mm^{-2} for FVIII light chain variants and 39 fmol/mm^{-2} for FVIIIdB variants) was immobilized onto a CM5 sensor chip using the amine coupling method according to the manufacturer's instructions. Subsequently, FVIII light chain variants were bound directly from medium to the anti-C2 antibody at a density of 17 fmol/mm^{-2} and purified FVIIIdB variants were bound at a density of 9 fmol/mm^{-2}. LRP1 cluster II (0.2-200 nM) was passed over the FVIII light chain or FVIIIdB variants in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 µL/min. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIII light chain or FVIIIdB variants was corrected for binding in absence of FVIII. Binding data during the association phase were fitted in a one-phase exponential association model. Responses at equilibrium were plotted as a function of the LRP1 cluster II concentration. The responses at equilibrium were fitted by non-linear regression using a standard hyperbola to generate K_D values assuming similar B_{max} values (GraphPad Prism 4 software, San Diego, CA, USA). To study the contribution of lysine residues on the interaction between LRP1 cluster II and the FVIII light chain, lysine residues were modified by passing over 50 mM of sulfo-NHS acetate or sulfo-NHS biotin (ThermoFisher Scientific, Rockford, IL, USA) for 10 min at 25 °C with a flow rate of 20 µL/min prior to passing over LRP1 cluster II (0.2-200 nM) and anti-*a3* antibody CLB-CAg69(39) (0.1-100 nM). Binding of purified FVIII light chain and FVIIIdB variants to full length LRP1 was studied by coupling LRP1 (Biomac, Leipzig, Germany) directly on a CM5 chip according to the manufacturer's instructions at three different densities (15 fmol/mm^{-2}, 18 fmol/mm^{-2} and 21 fmol/mm^{-2}). FVIII light chain variants (1-25 µg/ml based on Bradford analysis) and FVIIIdB variants (5-1000U/ml based on chromogenic activity) were passed over the immobilized LRP1 in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 µL/min. The sensor chip surface was regenerated three times after each concentration of FVIII using the same buffer containing 1 M NaCl. Binding to LRP1 was corrected for

binding in absence of LRP1. The response units at time point 235 seconds were plotted as a function of the concentration.

Structural modeling -The illustrative model of the triplet canonical binding mode of LRP to the light chain of FVIII is constructed using the Rosetta Modeling Suite(40) and the NAMD molecular dynamics software package(41). The basic preparation procedure was as follows, a detailed description can found in the Supplementary Materials section. Based on structural alignment with the x-ray crystallographic structure of the LDLR-RAP complex (PDB: 2FCW)(12), a homology model of the calcium-loaded LRP double complementary repeat (CR) 56 was created and aligned with the FVIII light chain (PDB: 4BDV, chain B)(42). Lysine residues (K2065 and K2092) were aligned to reconstruct the LDL family ligand receptor binding motif. A further CR5 module was positioned to interact with lysine K2136 of the FVIII light chain and was connected to its adjacent CR56 module by loop modeling. The resulting model was energy minimized prior to being equilibrated using molecular dynamics simulations for a short period of time, giving the structural model of the proposed canonical interaction between LRP and the FVIII light chain, as seen in Figure 6 (right).

RESULTS

HDX-MS of LRP1 interaction with cofactor FVIII -HDX mass spectrometry was employed to structurally elucidate the LRP1 binding surface of the FVIII light chain. HDX of FVIII in the presence and absence of LRP1 cluster II was followed at various time intervals ranging from 10 seconds to 256 minutes. Inline pepsin protein digest conditions were established that afforded peptides of reasonably small length (typically 5-20 amino acids). These conditions enabled detection of a total of 127 peptides covering 89% of the primary sequence of the FVIII light chain, with redundant coverage of all subunits, i.e., A3, C1, and C2 (Figure 1). The small areas of the N-terminus of *a3*, and of A3 contributed to 5% of the 11% gap in resolution. The majority of the peptides displayed an exchange pattern that is unaffected by the binding of LRP1 cluster II (Figure 1, blue color). However, several overlapping peptic peptide fragments encompassing residues 1671-1689, 1737-1746, 1804-1835, 2057-2069, 2078-2099, 2213-2235, 2266-2271, and 2293-2296 showed reduced HDX in the presence of LRP1 cluster II (Figure 1, red color). The raw data of representative HDX profiles for these peptides are shown in Figure 2. The unaffected peptide fragments (blue) and the peptides showing reduced HDX (red) were color coded in the crystal structure of FVIII (pdb code 4BDV) (Figure 2). This analysis revealed that the regions that demonstrate reduced HDX are scattered throughout the *a3*, A3, C1 and C2 domains of the FVIII light chain.

Chemical modification of lysine residues in the FVIII light chain abolishes the interaction with LRP1 cluster II -To address the role of lysine residues in the interaction between LRP1

cluster II and the FVIII light chain, we chemically modified lysine residues of the FVIII light chain on a SPR sensor chip using the compounds sulfo-NHS-acetate and sulfo-NHS-biotin. Both these compounds form stable, covalent amide bonds with primary amines, resulting in the replacement of the positively charged ε-amino group of the lysine residue by an uncharged acetate group (sulfo-NHS-acetate) or biotin group (sulfo-NHS-biotin). Subsequently, LRP1 cluster II and a control antibody CLB-CAg69 directed against the *a3* domain of FVIII, were passed over the unmodified and chemically modified FVIII light chain. While LRP1 cluster II readily bound to the unmodified FVIII light chain, no detectable binding was observed for the FVIII light chain in which the lysine residues were chemically modified (Figure 3). As a control, the CLB-CAg69 antibody bound to a similar extent to the unmodified as well as to the chemically modified FVIII light chain (Figure 3). These data suggest that lysine-containing surface patches mediate the interaction between LRP1 cluster II and the FVIII light chain.

Multiple lysine residues in the FVIII light chain contribute to the interaction with LRP1 cluster II

-To systematically address which lysine residues in the FVIII light chain contribute to LRP1 cluster II interaction, we constructed a library of FVIII variants carrying lysine to arginine (KR) replacements. All of these variants were tested for LRP1 cluster II interaction using SPR analysis. On each SPR sensor chip we included a control channel (only CLB-EL14 IgG4), wild type FVIII light chain and two FVIII light chain variants. Therefore, wild type FVIII light chain was analyzed multiple times (n=14). For each channel, binding data during the association phase were fitted in a one-phase exponential association model. A typical example of the binding curves and the fit of the one-phase exponential association model (dotted line) for wild type FVIII light chain are represented in Figure 4A. Responses at equilibrium were plotted as a function of the LRP1 cluster II concentration and fitted by non-linear regression using a standard hyperbola to generate K_D values (Supplemental Figure 1). The K_D value for wild type FVIII light chain was 30 ± 4 nM (Figure 4B). Initially, we constructed FVIII light chain variants based on the domain structure of FVIII carrying 2 (*a3*), 17 (A3), 8 (C1) or 9 (C2) replacements. However, KR variants of the A3 domain (17 replacements) and the C1 domain (8 replacements) were poorly expressed. We therefore constructed KR variants carrying single, double or triple lysine to arginine replacements in the A3 and C1 domain. Unfortunately, the K1732R FVIII light chain variant was poorly expressed and could not be assessed using SPR analysis. For all of the other variants, binding to LRP1 cluster II was analyzed and the K_D was calculated as described above (Figure 4C). The K_D for LRP1 cluster II for these variants was compared to the K_D of wild type (30 ± 4 nM) using a two-tailed students t-test. Our SPR analysis showed that lysine residues in the

C2 domain of the FVIII light chain do not contribute to the interaction with LRP1 cluster II. In contrast, multiple lysine residues in the *a3*A3C1 domains contributed to the interaction of FVIII to LRP1, although none of the individual substitutions completely abolished LRP1 cluster II binding. Lysine residues contributing to the interaction included K1673/1674, K1693, K1813/K1818, K1827, K1967, K1972, K2065, K2092 and K2136. Of these, K1693, K1827 and K1967 ranked the highest in the A3 domain, and K2065 and K2092 in the C1 domain. These residues were selected for further studies.

Single and multiple replacements of selected lysine residues in the A3 domain K1693, K1827, K1967 and in the C1 domain K2065 and K2092 by arginine, alanine or glutamic acid -From the SPR data of our library of FVIII light chain variants, we concluded that multiple lysine residues contribute to the interaction with LRP1 cluster II. However, there are some lysine residues in the A3 domain (K1693, K1827 and K1967) and in the C1 domain (K2065 and K2092) that contribute most to the interaction. To study the effect of charge, these lysine residues were replaced by positively charged arginine, uncharged alanine or negatively charged glutamic acid. The variant carrying a K1967E replacement was poorly expressed and could not be analyzed. All other FVIII light chain variants demonstrated a reduced interaction with LRP cluster II (Supplemental Figure 1 and Table 1), confirming the contribution of lysine residues at positions 1693, 1827, 1967, 2065 and 2092 to the interaction with LRP1 cluster II. There were no major differences in the interaction with LRP1 cluster II for FVIII light chain variants carrying lysine to arginine or lysine to alanine replacements for the positions 1967, 2065 and 2092. This confirms our previous observation that, despite their positive charge, arginine cannot replace lysine in the interaction with LRP1. However, for lysine residues at positions 1693 and 1827, alanine replacements demonstrated a higher K_D as compared to arginine replacements (Table 1), suggesting that although for these specific residues a lysine residue is preferred, the charge of the arginine can maintain some interaction with LRP1. In addition, we have reversed the positive charge of the lysine residue by replacing the lysine with a negatively charged glutamic acid residue. Alanine and glutamic acid variants demonstrated a similar K_D, indicating that there was no additive effect of neutralizing the charge (Table 1). To study the additive effect of combining lysine replacements, we constructed FVIII variants carrying 3 (K1693, K1827 and K1967), 2 (K2065 and K2092) and 5 lysine replacements. All FVIII light chain variants carrying the K1967E replacement were poorly expressed and could not be analyzed. All the other variants demonstrated a reduced interaction with LRP1 cluster II (Supplemental Figure 3 and Table 2). FVIII light chain variants containing 3 lysine replacements in the A3 domain demonstrated a slightly increased K_D as compared to the variants

containing only 1 lysine replacement in the A3 domain. There was no major difference between the lysine to arginine and lysine to alanine variants (Table 2). For the C1 domain, the K_D for variants in which both K2065 and K2092 were replaced was increased as compared to the C1 variants carrying only 1 lysine replacement. For the C1 domain, alanine or glutamic acid replacements did result in a higher K_D as compared to arginine replacements (Table 2). The combined variants in which the 5 lysine residues in the A3 and C1 were replaced by arginine or alanine demonstrated a K_D which was approximately 20-fold (arginine) or 40-fold (alanine) reduced as compared to wild type FVIII light chain (Table 2). In addition to interaction studies by capturing FVIII light chain variants directly from medium via the C2-domain on a SPR chip, interaction studies were performed using purified FVIII light chain variants and full length LRP1 (Supplemental Figure 4). These studies showed that, although the FVIII light chain variant carrying 5 lysine to arginine replacements demonstrated a 20-fold reduced interaction with LRP1 cluster II, binding to full length LRP1 was not affected. However, replacements at the same position by alanine or glutamic acid did result in a reduced interaction with full length LRP1 (Supplemental Figure 4).

From light chain to heterodimer: Binding of FVIIIdB variants to LRP1 cluster II and full length LRP1 -So far, we performed our binding assays using FVIII light chain variants. The FVIII light chain may have a different conformation as compared to the FVIII cofactor. To verify the contribution of the 5 lysine residues in the FVIII cofactor, we purified FVIIIdB variants carrying 5 lysine to arginine, alanine or glutamic acid replacements. FVIII variants were bound via the C2 domain to antibody CLB EL14 IgG4 and a concentration range of LRP1 cluster II was passed over the immobilized FVIII variants. These SPR binding studies indicated that all FVIIIdB variants showed a dramatically reduced interaction with LRP1 cluster II in the order wild type>>>5xKR>5xKA>>5xKE (Figure 5). While for individual lysine residues, there was only a minor difference between KR, KA and KE variants, the combined KE replacements proved to be more detrimental for the interaction with LRP1 cluster II than KA and KR replacements. In addition, we studied the contribution of the 5 lysine residues in the interaction with full length LRP1. Therefore, we directly coupled full length LRP1 on the CM5 chip and passed over the FVIIIdB variants. Binding studies indicated that the binding was reduced in the order 5xKR>5xKA>5xKE. The binding defect of the 5xKR variant was remarkably minor while the binding defects of the 5xKA and 5xKE were more severe (Supplemental Figure 5).

DISCUSSION

This study demonstrates that lysine residues indeed make a dominant contribution to the LRP1-FVIII interaction. When we chemically modify all surface exposed lysine residues of

FVIII using NHS-biotin or NHS-sulfo acetate, the binding to LRP1 cluster II is completely abrogated. Our combined experimental approach using HDX-MS and site-directed mutagenesis further suggests that the interaction of LRP1 with FVIII occurs over an extended surface containing multiple lysine residues in FVIII. Hydrogen deuterium exchange-mass spectrometry (HDX-MS) experiments identified regions 1671-1689, 1737-1746, 1804-1835, 2057-2069, 2078-2099, 2213-2235, 2266-2271, and 2293-2296 to be involved in the binding of FVIII to LRP1 cluster II. Intriguingly, these regions are scattered throughout the *a3*, A3, C1 and C2 domains of FVIII. Subsequent SPR analysis of a library of FVIII light chain variants carrying lysine to arginine (KR), alanine (KA) or glutamic acid (KE) revealed that multiple (11) lysine residues within the FVIII light chain including 1673/1674, 1693, 1813/1818, 1827, 1967, 1972, 2065, 2092 and 2136 contribute to the interaction with LRP1 cluster II. However, none of the individual lysine residues accounts completely for LRP1 cluster II binding. We did observe an additive effect of combining multiple lysine replacements and further mutagenesis studies using positively charged arginine, uncharged alanine or negatively charged glutamic acid residues indicated that reversing the positive charge had the strongest effect on the interaction.

The majority of the lysine residues that were identified in the mutagenesis study are located within those regions that were also identified using the HDX-MS study and are therefore considered 'hot spots'. These include 1673/1674, 1813/1818, 1827, 2065 and 2092 in FVIII. These data are in good agreement with previously described interactive sites on FVIII that mediate the binding to LRP1(21,35) and are fully compatible with our previous observation that the A3C1 domains of FVIII harbor important interaction sites for LRP1(43). Taken together, the pattern that emerges suggests that ligand binding to LRP1 is supported by a variety of target binding sites that most likely interact simultaneously with multiple CR repeats. The interaction of each individual site may be weak, however, when many interaction sites are in play at the same time dissociation of a single site will not allow abrogation of the complex. Hence, the combination of many weak binding sites results in an overall strong interaction, suggesting a cooperative model of interaction. This may explain our observations that the binding defects of both FVIII light chain and FVIIIdB variants containing 5xR replacements for full length LRP1 cluster II were remarkably minor compared to LRP1 cluster II. In contrast, the binding defects of the FVIII light chain and FVIIIdB variants containing 5xA and 5xE replacements were comparable between LRP1 cluster II and LRP1.

The question remains why certain lysine residues contribute more to the interaction than others. Based on a study on the specificity of binding of the serpins PAI-1 and PN-1 to LRP1(30), Jensen *et al.* have suggested that there may be modest requirements to be a coordinating

lysine residue. However, inspection of the x-ray structure of the LDLR-RAP complex suggests that members of the LDL receptor family interact with clusters of lysines being 18-20 Å apart(12). In addition, it has been suggested that hydrophobic interactions may play an additional role in the binding of ligands to the complement-type repeats(14). While the acidic necklace model as proposed by Fisher *et al.* suggests that hydrophobic interactions are mediated by the side chain of the lysine residue of the ligand(12), Jensen *et al.* propose that an additional hydrophobic residue of the ligand, typically a leucine or an isoleucine, is important for the interaction(14). Therefore, hydrophobic residues located in the vicinity of the lysine residue may determine whether a lysine residue is a 'hot-spot' or not. Alternatively, the presence of a second proximal charge near the lysine residue, or perhaps even repulsive charges from amino acid residues surrounding the lysine residue, may determine specificity(17,29).

We therefore took a close look at the spatial distribution of the identified interactive regions and lysine residues within FVIII. Crystal structure analysis revealed that all 'hot spot' lysine residues that we identified may spatially align with the acidic binding pockets of the extracellular part of the LDL receptor. The distance from the bottom of the C1 domain to the top of the A3 domain spans 69 Å, a distance that LRP1 cluster II can easily encompass using just 3-4 complement type repeats. We therefore propose that LRP1 cluster II interacts with the FVIII light chain via an extended surface that starts at the bottom of the C1 domain and extends to the top of the A3 domain. This prompted us to perform structural modeling with the goal to both exemplify and substantiate likely interactions between LRP1 cluster II and the FVIII light chain. Initially, we focused on the FVIII C1 domain in which both HDX-MS and mutagenesis studies indicated the contribution of 2 lysine residues in the C1 domain, K2065 and K2092. Homology modeling was performed in consensus with the canonical binding mode for members of the LDL receptor family using the X-ray crystallographic structure of the LDLR-RAP complex as a template(12). Hence, each CR module contains three acidic aspartic acid residues placed in a concave surface region which interacts with lysine clusters. The proposed canonical interaction mode between LRP1 triple module and the FVIII light chain allows two principal binding modes oppositely oriented relative to each other, i.e. LRP1 N-terminus to C-terminus aligns along the K2065-K2092-K2136 as opposed to K2136-K2092-K2065 surface patch. Two key features contribute to a favorable interaction between lysine regions and the LRP triple module. Shape complementarity is eminently achievable due to flexible nature of the chain of complementary repeats allowing the receptor to adapt to the ligands surface topology and exploit hydrophobic interaction in between repeats. Also, the ionic interactions between the positively charged lysine residues and the negatively charged

aspartic acid residues resulted in close packing of the partners suggesting that these features are the driving forces behind the interaction. The wide spread imprint of LRP1 binding on FVIII identified by HDX-MS analysis points to the notion that during LRP1 binding CR repeats switch between different lysine clusters. This is reflected in the designed LRP1 model where several interacting modes are equally likely.

In addition to the overlap of HDX and the lysine mutagenesis study, there were also some differences. Residues 1693, 1967, 1972 and 2136 were identified only in the lysine mutagenesis study, while regions 1737-1746, 2213-2235, 2266-2271 and 2293-2296 were only identified using the HDX-MS approach. We therefore consider these regions and lysine residues 'soft-spots'. The effect of the 'soft-spot' lysine residues may be explained by conformational changes; lysine residue 1693 is directly facing the 'hot-spot' lysine at position 1818 and the soft-spot lysine residue 1967 has been shown to interact with the A2 domain of FVIII(44). As for the 'soft-spot' regions identified by HDX-MS: region 1737-1746 is located within the A3 region and consists of the amino acids EFTDGSFTQP. The remaining 3 peptides are located within the C2 domain. The C2 domain has previously been reported to bind to LRP1 with low affinity ($K_D \approx$ 3.6 μM)(43). The peptide 2293-2296 (VVNS) contains no positively charged residues, the peptide 2266-2271 (QDGHQW) contains 1 histidine residue and peptide 2213-2235 (QGRSNAWRPQVNNPKEWLQVDFQ) contains 2 arginine residues and 1 lysine residue. We did not observe any contribution of lysine residues in the C2 domain of FVIII to the interaction with LRP1 in our mutagenesis study (Figure 4C), suggesting that the reported low affinity of the C2 domain may be mediated by residues other than lysine.

In recent years the importance of the C domains in the life cycle of FVIII has been appreciated since they are involved in lipid binding, VWF interaction and cellular uptake(45-50). The C2 domain is loosely attached to the C1 and A1 domains and it is likely that the domain has a high degree of rotational freedom. We believe that the positioning of the FVIII C domains relative to each other may therefore be an important aspect in the biology of FVIII. The regions in the C2 domain that were identified by HDX-MS may reflect conformational changes within these areas as opposed to a direct interaction with LRP1 cluster II. It would therefore be highly interesting to study the dynamics of the FVIII C domains in response to interaction with various ligands and lipids using HDX-MS technology.

REFERENCES

1. Herz, J., Chen, Y., Masiulis, I., and Zhou, L. (2009) Expanding functions of lipoprotein receptors. *Journal of lipid research* **50 Suppl**, S287-292
2. Krieger, M., and Herz, J. (1994) Structures and functions of multiligand lipoprotein receptors: macrophage scavenger receptors and LDL receptor-related protein (LRP). *Annual review of biochemistry* **63**, 601-637
3. Dieckmann, M., Dietrich, M. F., and Herz, J. (2010) Lipoprotein receptors--an evolutionarily ancient multifunctional receptor family. *Biological chemistry* **391**, 1341-1363
4. Herz, J., and Strickland, D. K. (2001) LRP: a multifunctional scavenger and signaling receptor. *The Journal of clinical investigation* **108**, 779-784
5. Strickland, D. K., and Ranganathan, S. (2003) Diverse role of LDL receptor-related protein in the clearance of proteases and in signaling. *Journal of thrombosis and haemostasis : JTH* **1**, 1663-1670
6. Fass, D., Blacklow, S., Kim, P. S., and Berger, J. M. (1997) Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. *Nature* **388**, 691-693
7. Andersen, O. M., Christensen, L. L., Christensen, P. A., Sorensen, E. S., Jacobsen, C., Moestrup, S. K., Etzerodt, M., and Thogersen, H. C. (2000) Identification of the minimal functional unit in the low density lipoprotein receptor-related protein for binding the receptor-associated protein (RAP). A conserved acidic residue in the complement-type repeats is important for recognition of RAP. *The Journal of biological chemistry* **275**, 21017-21024
8. Brown, M. S., Herz, J., and Goldstein, J. L. (1997) LDL-receptor structure. Calcium cages, acid baths and recycling receptors. *Nature* **388**, 629-630
9. Moestrup, S. K., Kaltoft, K., Sottrup-Jensen, L., and Gliemann, J. (1990) The human alpha 2-macroglobulin receptor contains high affinity calcium binding sites important for receptor conformation and ligand recognition. *The Journal of biological chemistry* **265**, 12623-12628
10. Strickland, D. K., Ashcom, J. D., Williams, S., Battey, F., Behre, E., McTigue, K., Battey, J. F., and Argraves, W. S. (1991) Primary structure of alpha 2-macroglobulin receptor-associated protein. Human homologue of a Heymann nephritis antigen. *The Journal of biological chemistry* **266**, 13364-13369
11. Herz, J., Goldstein, J. L., Strickland, D. K., Ho, Y. K., and Brown, M. S. (1991) 39-kDa protein modulates binding of ligands to low density lipoprotein receptor-related protein/alpha 2-macroglobulin receptor. *The Journal of biological chemistry* **266**, 21232-21238
12. Fisher, C., Beglova, N., and Blacklow, S. C. (2006) Structure of an LDLR-RAP complex reveals a general mode for ligand recognition by lipoprotein receptors. *Molecular cell* **22**, 277-283
13. Migliorini, M. M., Behre, E. H., Brew, S., Ingham, K. C., and Strickland, D. K. (2003) Allosteric modulation of ligand binding to low density lipoprotein receptor-related protein by the receptor-associated protein requires critical lysine residues within its carboxyl-terminal domain. *The Journal of biological chemistry* **278**, 17986-17992
14. Jensen, G. A., Andersen, O. M., Bonvin, A. M., Bjerrum-Bohr, I., Etzerodt, M., Thogersen, H. C., O'Shea, C., Poulsen, F. M., and Kragelund, B. B. (2006) Binding

site structure of one LRP-RAP complex: implications for a common ligand-receptor binding motif. *Journal of molecular biology* **362**, 700-716
15. van den Biggelaar, M., Sellink, E., Klein Gebbinck, J. W., Mertens, K., and Meijer, A. B. (2011) A single lysine of the two-lysine recognition motif of the D3 domain of receptor-associated protein is sufficient to mediate endocytosis by low-density lipoprotein receptor-related protein. *The international journal of biochemistry & cell biology* **43**, 431-440
16. Zaiou, M., Arnold, K. S., Newhouse, Y. M., Innerarity, T. L., Weisgraber, K. H., Segall, M. L., Phillips, M. C., and Lund-Katz, S. (2000) Apolipoprotein E;-low density lipoprotein receptor interaction. Influences of basic residue and amphipathic alpha-helix organization in the ligand. *Journal of lipid research* **41**, 1087-1095
17. Dolmer, K., Campos, A., and Gettins, P. G. (2013) Quantitative dissection of the binding contributions of ligand lysines of the receptor-associated protein (RAP) to the low density lipoprotein receptor-related protein (LRP1). *The Journal of biological chemistry* **288**, 24081-24090
18. Bu, G., Geuze, H. J., Strous, G. J., and Schwartz, A. L. (1995) 39 kDa receptor-associated protein is an ER resident protein and molecular chaperone for LDL receptor-related protein. *The EMBO journal* **14**, 2269-2280
19. Willnow, T. E., Armstrong, S. A., Hammer, R. E., and Herz, J. (1995) Functional expression of low density lipoprotein receptor-related protein is controlled by receptor-associated protein in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 4537-4541
20. Willnow, T. E., Rohlmann, A., Horton, J., Otani, H., Braun, J. R., Hammer, R. E., and Herz, J. (1996) RAP, a specialized chaperone, prevents ligand-induced ER retention and degradation of LDL receptor-related endocytic receptors. *The EMBO journal* **15**, 2632-2639
21. Meems, H., van den Biggelaar, M., Rondaij, M., van der Zwaan, C., Mertens, K., and Meijer, A. B. (2011) C1 domain residues Lys 2092 and Phe 2093 are of major importance for the endocytic uptake of coagulation factor VIII. *The international journal of biochemistry & cell biology* **43**, 1114-1121
22. Howard, G. C., Yamaguchi, Y., Misra, U. K., Gawdi, G., Nelsen, A., DeCamp, D. L., and Pizzo, S. V. (1996) Selective mutations in cloned and expressed alpha-macroglobulin receptor binding fragment alter binding to either the alpha2-macroglobulin signaling receptor or the low density lipoprotein receptor-related protein/alpha2-macroglobulin receptor. *The Journal of biological chemistry* **271**, 14105-14111
23. Lalazar, A., Weisgraber, K. H., Rall, S. C., Jr., Giladi, H., Innerarity, T. L., Levanon, A. Z., Boyles, J. K., Amit, B., Gorecki, M., Mahley, R. W., and et al. (1988) Site-specific mutagenesis of human apolipoprotein E. Receptor binding activity of variants with single amino acid substitutions. *The Journal of biological chemistry* **263**, 3542-3545
24. Nielsen, K. L., Holtet, T. L., Etzerodt, M., Moestrup, S. K., Gliemann, J., Sottrup-Jensen, L., and Thogersen, H. C. (1996) Identification of residues in alpha-macroglobulins important for binding to the alpha2-macroglobulin receptor/Low density lipoprotein receptor-related protein. *The Journal of biological chemistry* **271**, 12909-12912

25. Rodenburg, K. W., Kjoller, L., Petersen, H. H., and Andreasen, P. A. (1998) Binding of urokinase-type plasminogen activator-plasminogen activator inhibitor-1 complex to the endocytosis receptors alpha2-macroglobulin receptor/low-density lipoprotein receptor-related protein and very-low-density lipoprotein receptor involves basic residues in the inhibitor. *The Biochemical journal* **329 (Pt 1)**, 55-63
26. Stefansson, S., Muhammad, S., Cheng, X. F., Battey, F. D., Strickland, D. K., and Lawrence, D. A. (1998) Plasminogen activator inhibitor-1 contains a cryptic high affinity binding site for the low density lipoprotein receptor-related protein. *The Journal of biological chemistry* **273**, 6358-6366
27. Williams, S. E., Inoue, I., Tran, H., Fry, G. L., Pladet, M. W., Iverius, P. H., Lalouel, J. M., Chappell, D. A., and Strickland, D. K. (1994) The carboxyl-terminal domain of lipoprotein lipase binds to the low density lipoprotein receptor-related protein/alpha 2-macroglobulin receptor (LRP) and mediates binding of normal very low density lipoproteins to LRP. *The Journal of biological chemistry* **269**, 8653-8658
28. Lee, C. J., De Biasio, A., and Beglova, N. (2010) Mode of interaction between beta2GPI and lipoprotein receptors suggests mutually exclusive binding of beta2GPI to the receptors and anionic phospholipids. *Structure* **18**, 366-376
29. Gettins, P. G., and Dolmer, K. (2012) A proximal pair of positive charges provides the dominant ligand-binding contribution to complement-like domains from the LRP (low-density lipoprotein receptor-related protein). *The Biochemical journal* **443**, 65-73
30. Jensen, J. K., Dolmer, K., and Gettins, P. G. (2009) Specificity of binding of the low density lipoprotein receptor-related protein to different conformational states of the clade E serpins plasminogen activator inhibitor-1 and proteinase nexin-1. *The Journal of biological chemistry* **284**, 17989-17997
31. van den Biggelaar, M., Bierings, R., Storm, G., Voorberg, J., and Mertens, K. (2007) Requirements for cellular co-trafficking of factor VIII and von Willebrand factor to Weibel-Palade bodies. *Journal of thrombosis and haemostasis : JTH* **5**, 2235-2242
32. van den Biggelaar, M., Meijer, A. B., Voorberg, J., and Mertens, K. (2009) Intracellular cotrafficking of factor VIII and von Willebrand factor type 2N variants to storage organelles. *Blood* **113**, 3102-3109
33. Thim, L., Vandahl, B., Karlsson, J., Klausen, N. K., Pedersen, J., Krogh, T. N., Kjalke, M., Petersen, J. M., Johnsen, L. B., Bolt, G., Norby, P. L., and Steenstrup, T. D. (2010) Purification and characterization of a new recombinant factor VIII (N8). *Haemophilia : the official journal of the World Federation of Hemophilia* **16**, 349-359
34. Fribourg, C., Meijer, A. B., and Mertens, K. (2006) The interface between the EGF2 domain and the protease domain in blood coagulation factor IX contributes to factor VIII binding and factor X activation. *Biochemistry* **45**, 10777-10785
35. Bovenschen, N., Boertjes, R. C., van Stempvoort, G., Voorberg, J., Lenting, P. J., Meijer, A. B., and Mertens, K. (2003) Low density lipoprotein receptor-related protein and factor IXa share structural requirements for binding to the A3 domain of coagulation factor VIII. *The Journal of biological chemistry* **278**, 9370-9377
36. Andersen MD, F. J. (2011) Structural characterization of both the non-proteolytic and proteolytic activation pathways of coagulation Factor XIII studied by

hydrogen-deuterium exchange mass spectrometry. *Int J Mass Spectrom* **302**, 139-148
37. Kavan, D. M., P. (2011) MSTools - Web based application for visualization and presentation of HXMS data. *Int. J. Mass Spectrom.* **302**, 53-58
38. van Helden, P. M., van den Berg, H. M., Gouw, S. C., Kaijen, P. H., Zuurveld, M. G., Mauser-Bunschoten, E. P., Aalberse, R. C., Vidarsson, G., and Voorberg, J. (2008) IgG subclasses of anti-FVIII antibodies during immune tolerance induction in patients with hemophilia A. *British journal of haematology* **142**, 644-652
39. Leyte, A., Mertens, K., Distel, B., Evers, R. F., De Keyzer-Nellen, M. J., Groenen-Van Dooren, M. M., De Bruin, J., Pannekoek, H., Van Mourik, J. A., and Verbeet, M. P. (1989) Inhibition of human coagulation factor VIII by monoclonal antibodies. Mapping of functional epitopes with the use of recombinant factor VIII fragments. *The Biochemical journal* **263**, 187-194
40. Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y. E., Fleishman, S. J., Corn, J. E., Kim, D. E., Lyskov, S., Berrondo, M., Mentzer, S., Popovic, Z., Havranek, J. J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology* **487**, 545-574
41. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781-1802
42. Svensson, L. A., Thim, L., Olsen, O. H., and Nicolaisen, E. M. (2013) Evaluation of the metal binding sites in a recombinant coagulation factor VIII identifies two sites with unique metal binding properties. *Biological chemistry* **394**, 761-765
43. Lenting, P. J., Neels, J. G., van den Berg, B. M., Clijsters, P. P., Meijerman, D. W., Pannekoek, H., van Mourik, J. A., Mertens, K., and van Zonneveld, A. J. (1999) The light chain of factor VIII comprises a binding site for low density lipoprotein receptor-related protein. *The Journal of biological chemistry* **274**, 23734-23739
44. Bloem, E., Meems, H., van den Biggelaar, M., van der Zwaan, C., Mertens, K., and Meijer, A. B. (2012) Mass spectrometry-assisted study reveals that lysine residues 1967 and 1968 have opposite contribution to stability of activated factor VIII. *The Journal of biological chemistry* **287**, 5775-5783
45. Gilbert, G. E., Novakovic, V. A., Kaufman, R. J., Miao, H., and Pipe, S. W. (2012) Conservative mutations in the C2 domains of factor VIII and factor V alter phospholipid binding and cofactor activity. *Blood* **120**, 1923-1932
46. Meems, H., Meijer, A. B., Cullinan, D. B., Mertens, K., and Gilbert, G. E. (2009) Factor VIII C1 domain residues Lys 2092 and Phe 2093 contribute to membrane binding and cofactor activity. *Blood* **114**, 3938-3946
47. Gilbert, G. E., Kaufman, R. J., Arena, A. A., Miao, H., and Pipe, S. W. (2002) Four hydrophobic amino acids of the factor VIII C2 domain are constituents of both the membrane-binding and von Willebrand factor-binding motifs. *The Journal of biological chemistry* **277**, 6374-6381
48. Lu, J., Pipe, S. W., Miao, H., Jacquemin, M., and Gilbert, G. E. (2011) A membrane-interactive surface on the factor VIII C1 domain cooperates with the C2 domain for cofactor function. *Blood* **117**, 3181-3189

49. Bloem, E., van den Biggelaar, M., Wroblewska, A., Voorberg, J., Faber, J. H., Kjalke, M., Stennicke, H. R., Mertens, K., and Meijer, A. B. (2013) Factor VIII C1 domain spikes 2092-2093 and 2158-2159 comprise regions that modulate cofactor function and cellular uptake. *The Journal of biological chemistry* **288**, 29670-29679
50. Herczenik, E., van Haren, S. D., Wroblewska, A., Kaijen, P., van den Biggelaar, M., Meijer, A. B., Martinez-Pomares, L., ten Brinke, A., and Voorberg, J. (2012) Uptake of blood coagulation factor VIII by dendritic cells is mediated via its C1 domain. *The Journal of allergy and clinical immunology* **129**, 501-509, 509 e501-505
51. Chivian, D., and Baker, D. (2006) Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. *Nucleic acids research* **34**, e112
52. Ebina, T., Toh, H., and Kuroda, Y. (2009) Loop-length-dependent SVM prediction of domain linkers for high-throughput structural proteomics. *Biopolymers* **92**, 1-8

FIGURE LEGENDS

Table 1. Effect of replacement of lysine residues by positively charged arginine, uncharged alanine or negatively charged glutamic acid on the interaction between the FVIII light chain and LRP1 cluster II.

Association and dissociation of LRP1 cluster II to FVIII light chain variants K1693, K1827, K1967, K2065 and K2092 was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). The anti-C2 antibody CLB-EL14 IgG4 (26 fmol/mm^{-2}) was immobilized onto a CM5 sensor chip using the amine coupling method according to the manufacturer's instructions. Subsequently, FVIII light chain variants were bound to the anti-C2 antibody at a density of 17 fmol/mm^{-2}. LRP1 cluster II (0.2-200 nM) was passed over the FVIII light chain variants in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIII light chain variants was corrected for binding in absence of FVIII. Binding data during the association phase were fitted in a one-phase exponential association model.

Table 2. Effect of combining lysine replacements on the interaction between the FVIII light chain and LRP1 cluster II.

Association and dissociation of LRP1 cluster II to FVIII light chain variants carrying replacements at positions K1693, K1827 and K1967 (A3), K2065 and K2092 (C1) and K1693, K1827, K1967, K2065 and K2092 (A3C1) was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). The anti-C2 antibody CLB-EL14 IgG4 (26 fmol/mm^{-2}) was immobilized onto a CM5 sensor chip using the amine coupling method according to the manufacturer's instructions. Subsequently, FVIII light chain variants were bound to the anti-C2 antibody at a density of 17 fmol/mm^{-2}. LRP1 cluster II (0.2-200 nM) was passed over the FVIII light chain variants in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIII light chain variants was corrected for binding in absence of FVIII. Binding data during the association phase were fitted in a one-phase exponential association model.

Figure 1. HDX of FVIII and FVIII in complex with LRP1 monitored by mass spectrometry.

Sequence coverage of peptide fragments of LC FVIII. Peptides that display reduced deuterium incorporation (>2 S:D.) following complex formation with LRP1 are colored in *red* peptides that displayed no changes are colored in *blue*. Localization of exchanging amides in both N- and C-terminally overlapping peptides was performed by substraction, assuming complete back exchange of the two N-terminal hydrogen in each peptide. Domains of LC FVIII are shown schematically.

Figure 2. HDX of FVIII and FVIII incubated with LRP1 monitored by mass spectrometry. Surface plot of FVIII based on the crystal structure of FVIII (Protein Data Bank code 4BDV). Peptides that display reduced deuterium incorporation (>2 S:D.) following complex formation with LRP1 are colored in *red* peptides that displayed no changes are colored in *blue*. Areas uncovered by the peptide fragment are grey colored. Insert fitted HDX curves of the representative peptides of FVIII (blue colored) and FVIII incubated with LRP1 (red colored) are shown, with raw data indicated by solid squares and open circles.

Figure 3. Chemical modification of lysine residues in the FVIII light chain abolishes the interaction with LRP1 cluster II.

Association and dissociation of LRP1 cluster II to FVIII light chain was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). The anti-C2 antibody CLB-EL14 IgG4 (26 fmol/mm^{-2}) was immobilized onto a CM5 sensor chip using the amine coupling method according to the manufacturer's instructions. Subsequently, FVIII light chain was bound to the anti-C2 antibody at a density of 17 fmol/mm^{-2}. To study the contribution of lysine residues on the interaction between LRP1 cluster II and the FVIII light chain, lysine residues were modified by passing over 50 mM of sulfo-NHS acetate or sulfo-NHS biotin (ThermoFisher Scientific, Rockford, IL, USA) for 10 min at 25 °C with a flow rate of 20 μL/min. **A.** LRP1 cluster II (0.2-200 nM) and **B.** anti-*a3* antibody CLB-CAg69 (0.1-100 nM) were passed over the FVIII light chain in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIII light chain was corrected for binding in absence of FVIII. The response at 235 seconds after association is plotted as a function of the concentration.

Figure 4. Library of lysine to arginine replacements in the FVIII light chain identifies the contribution of multiple lysine residues in the interaction with LRP1 cluster II.

Association and dissociation of LRP1 cluster II to FVIII light chain variants was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB,

Uppsala, Sweden). The anti-C2 antibody CLB-EL14 IgG4 (26 fmol/mm^{-2}) was immobilized onto a CM5 sensor chip using the amine coupling method according to the manufacturer's instructions. Subsequently, FVIII light chain variants were bound to the anti-C2 antibody at a density of 17 fmol/mm^{-2}. LRP1 cluster II (0.2-200 nM) was passed over the FVIII light chain variants in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIII light chain variants was corrected for binding in absence of FVIII. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIII light chain variants was corrected for binding in absence of FVIII. Binding data during the association phase were fitted in a one-phase exponential association model. **A.** Representative experiment for the interaction between LRP1 cluster II and wild type FVIII light chain. **B.** On each SPR sensor chip we included a control channel (only CLB-EL14 IgG4), wild type FVIII light chain and two FVIII light chain variants. Therefore, wild type FVIII light chain was analyzed multiple times (n=14) **C.** The K_D for LRP1 cluster II for the FVIII light chain variants was compared to the K_D of wild type FVIII light chain (n=14 degrees of freedom=13 $p<^{*}0.10$ t-value=2.16, $p^{**}<0.05$ t-value=3.012, $p^{***}<0.001$ t-value=4.22) using a two-tailed students t-test.

Figure 5. Binding of FVIIIdB variants to LRP1 cluster II.

A. Association and dissociation of LRP1 cluster II to FVIIIdB variants was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). The anti-C2 antibody CLB-EL14 IgG4 (39 fmol/mm^{-2} for FVIIIdB variants) was immobilized onto a CM5 sensor chip using the amine coupling method according to the manufacturer's instructions. Subsequently, FVIIIdB variants were bound to the anti-C2 antibody at a density of 9 fmol/mm^{-2}. LRP1 cluster II (0-200 nM) was passed over the FVIIIdB variants in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIIIdB variants was corrected for binding in absence of FVIII. Binding data during the association phase were fitted in a one-phase exponential association model. **B.** The responses at equilibrium were fitted by non-linear regression using a standard hyperbola to generate K_D values (GraphPad Prism 4 software, San Diego, CA, USA).

Figure 6. Left: Schematic depiction of the suggested LRP1 cluster II interaction with the FVIII light chain with the canonical (triplet) mode emphasized. Right: Cartoon representation of the binding motif between the calcium-loaded LRP triple module (green) and FVIII C1 (gray). Aspartic acid residues of LRP and the "hot spot" lysine residues of FVIII are shown in sticks.

Supplemental Figure 1.

Association and dissociation of LRP1 cluster II to FVIII light chain variants was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). The anti-C2 antibody CLB-EL14 IgG4 (26 fmol/mm^{-2}) was immobilized onto a CM5 sensor chip using the amine coupling method according to the manufacturer's instructions. Subsequently, FVIII light chain variants were bound to the anti-C2 antibody at a density of 17 fmol/mm^{-2}.LRP1 cluster II (0.2-200 nM) was passed over the FVIII light chain variants in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIII light chain variants was corrected for binding in absence of FVIII. Binding data during the association phase were fitted in a one-phase exponential association model.

Supplemental Figure 2.

Association and dissociation of LRP1 cluster II to FVIII light chain variants K1693, K1827, K1967, K2065 and K2092 was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). The anti-C2 antibody CLB-EL14 IgG4 (26 fmol/mm^{-2}) was immobilized onto a CM5 sensor chip using the amine coupling method according to the manufacturer's instructions. Subsequently, FVIII light chain variants were bound to the anti-C2 antibody at a density of 17 fmol/mm^{-2}. LRP1 cluster II (0.2-200 nM) was passed over the FVIII light chain variants in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIII light chain variants was corrected for binding in absence of FVIII. Binding data during the association phase were fitted in a one-phase exponential association model.

Supplemental Figure 3.

Association and dissociation of LRP1 cluster II to FVIII light chain variants carrying replacements at positions K1693, K1827 and K1967 (A3), K2065 and K2092 (C1) and K1693, K1827, K1967,

K2065 and K2092 (A3C1) was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). The anti-C2 antibody CLB-EL14 IgG4 (26 fmol/mm^{-2}) was immobilized onto a CM5 sensor chip using the amine coupling method according to the manufacturer's instructions. Subsequently, FVIII light chain variants were bound to the anti-C2 antibody at a density of 17 fmol/mm^{-2}. LRP1 cluster II (0.2-200 nM) was passed over the FVIII light chain variants in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of LRP1 cluster II using the same buffer containing 1 M NaCl. Binding to FVIII light chain variants was corrected for binding in absence of FVIII. Binding data during the association phase were fitted in a one-phase exponential association model.

Supplemental Figure 4. Binding of FVIII light chain variants to full length LRP1.

A. Association and dissociation of FVIII light chain variants to full length LRP1 was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). LRP1 (Biomac, Leipzig, Germany) was coupled directly on a CM5 chip according to the manufacturer's instructions at three different densities (15 fmol/mm^{-2}, 18 fmol/mm^{-2} and 21 fmol/mm^{-2}). FVIII light chain variants (1-25 μg/ml based on Bradford analysis) were passed over the immobilized LRP1 in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of FVIII using the same buffer containing 1 M NaCl. Binding to LRP1 was corrected for binding in absence of LRP1. Shown are the SPR curves for the channel on which 21 fmol/mm^{-2} LRP1 was coupled. **B.** The response units at time point 235 seconds were plotted as a function of the concentration.

Supplemental Figure 5. Binding of FVIIIdB variants to full length LRP1.

A. Association and dissociation of FVIIIdB variants to full length LRP1 was assessed by Surface Plasmon Resonance (SPR) analysis employing a BIAcore 3000 biosensor (Biacore AB, Uppsala, Sweden). LRP1 (Biomac, Leipzig, Germany) was coupled directly on a CM5 chip according to the manufacturer's instructions at three different densities (15 fmol/mm^{-2}, 18 fmol/mm^{-2} and 21 fmol/mm^{-2}). FVIIIdB variants (5-1000U/ml based on chromogenic activity) were passed over the immobilized LRP1 in a buffer containing 150 mM NaCl, 5 mM CaCl2, 0.005% (v/v) Tween 20 and 20 mM Hepes (pH 7.4) at 25 °C with a flow rate of 20 μL/min. The sensor chip surface was regenerated three times after each concentration of FVIII using the same buffer containing 1 M NaCl. Binding to LRP1 was corrected for binding in absence of LRP1. Shown are the SPR curves

for the channel on which 21 fmol/mm^{-2} LRP1 was coupled. **B.** The response units at time point 235 seconds were plotted as a function of the concentration.

Table 1.

	arginine	alanine	glutamic acid
1693	63 ± 5	170 ± 19	154 ± 17
1827	80 ± 7	157 ± 16	130 ± 13
1967	73 ± 6	91 ± 5	-
2065	168 ±14	202 ± 17	210 ± 22
2092	72 ± 6	45 ± 4	98 ± 10

Table 2.

	arginine	alanine	glutamic acid
A3	201 ± 18	222 ± 20	-
C1	220 ± 21	620 ± 59	534 ± 52
A3C1	567 ± 48	1339 ± 144	-

Figure 1.

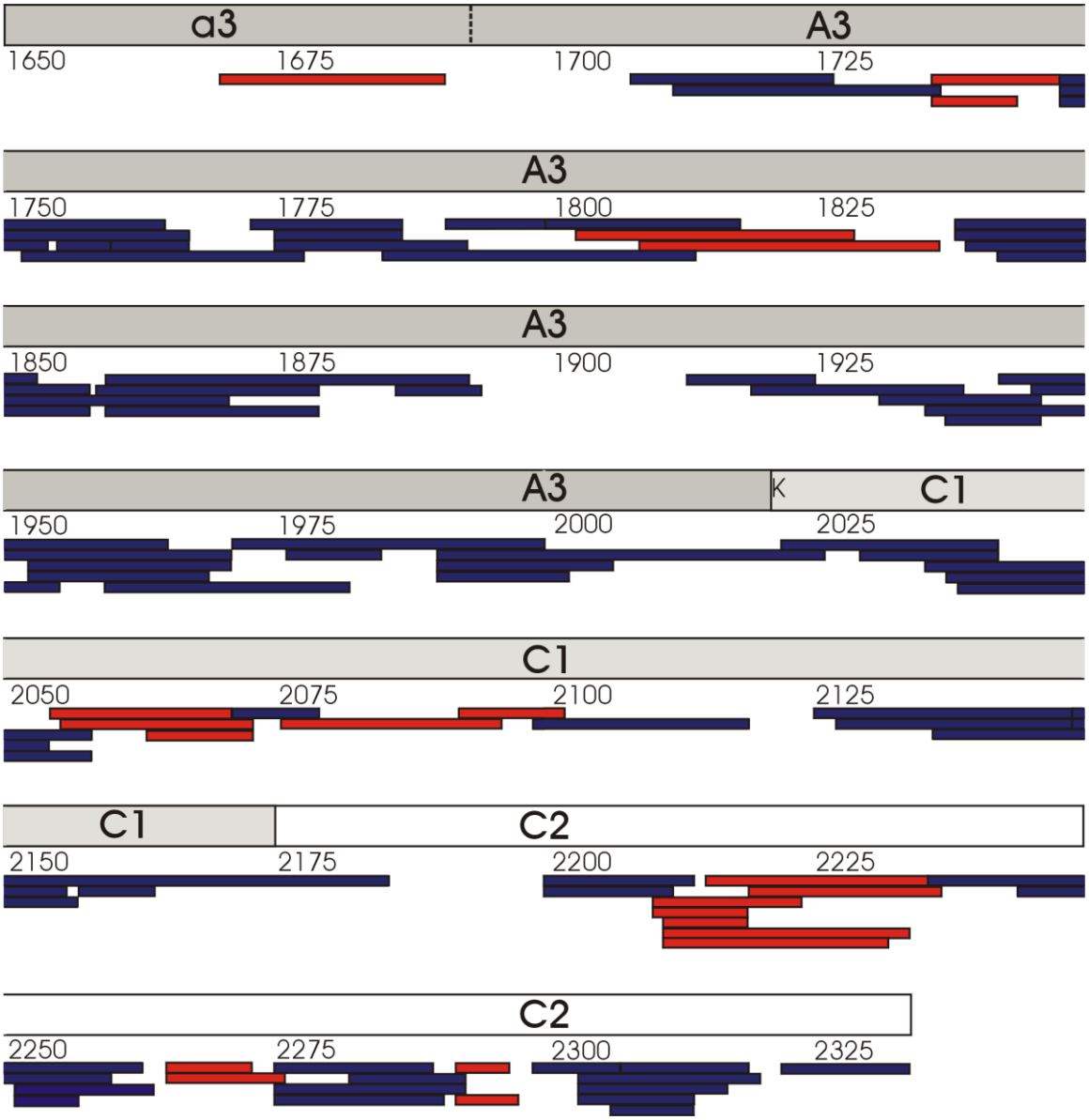

a3
A3
1650
1675
1700
1725
A3
1750
1775
1800
1825
A3
1850
1875
1900
1925
A3
K
C1
1950
1975
2000
2025
C1
2050
2075
2100
2125
C1
C2
2150
2175
2200
2225
C2
2250
2275
2300
2325

Figure 2.

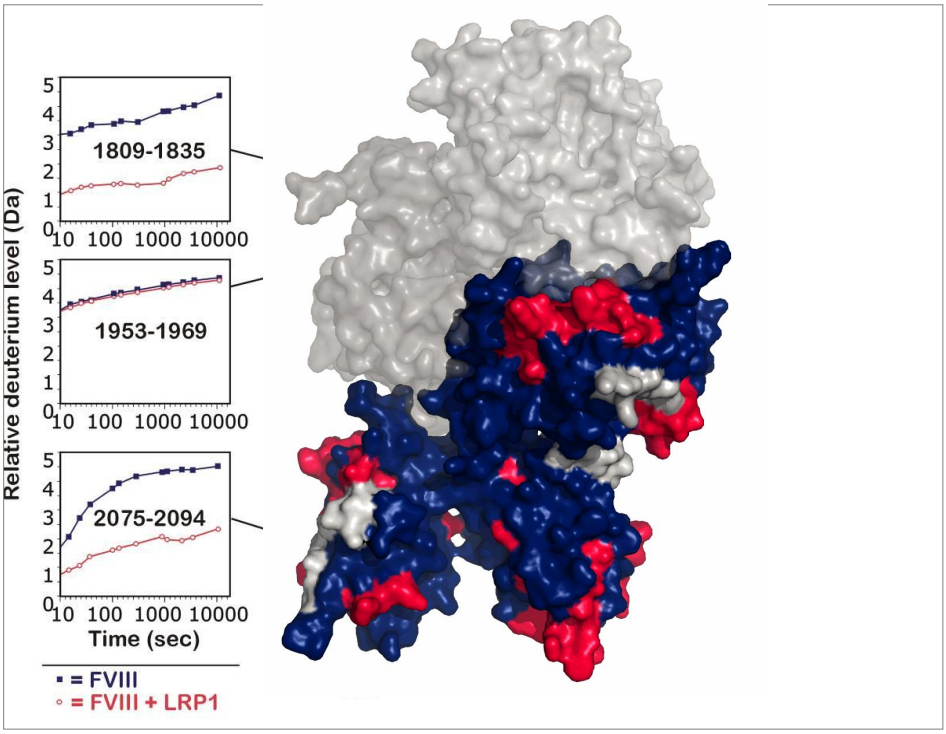

Relative deuterium level (Da)
1809-1835
1953-1969
2075-2094
10 100 1000 10000
Time (sec)
▪ = FVIII
◦ = FVIII + LRP1

Figure 3.

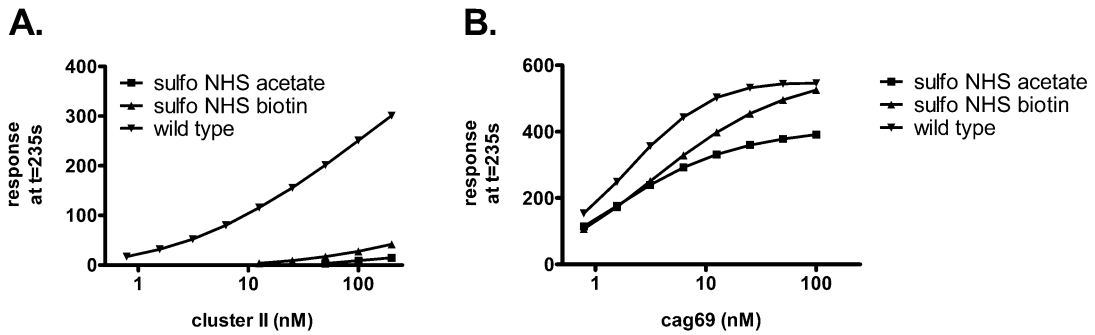

A.
B.
response at t=235s
sulfo NHS acetate
sulfo NHS biotin
wild type
cluster II (nM)
cag69 (nM)
0
100
200
300
400
600
1
10
100

Figure 4.

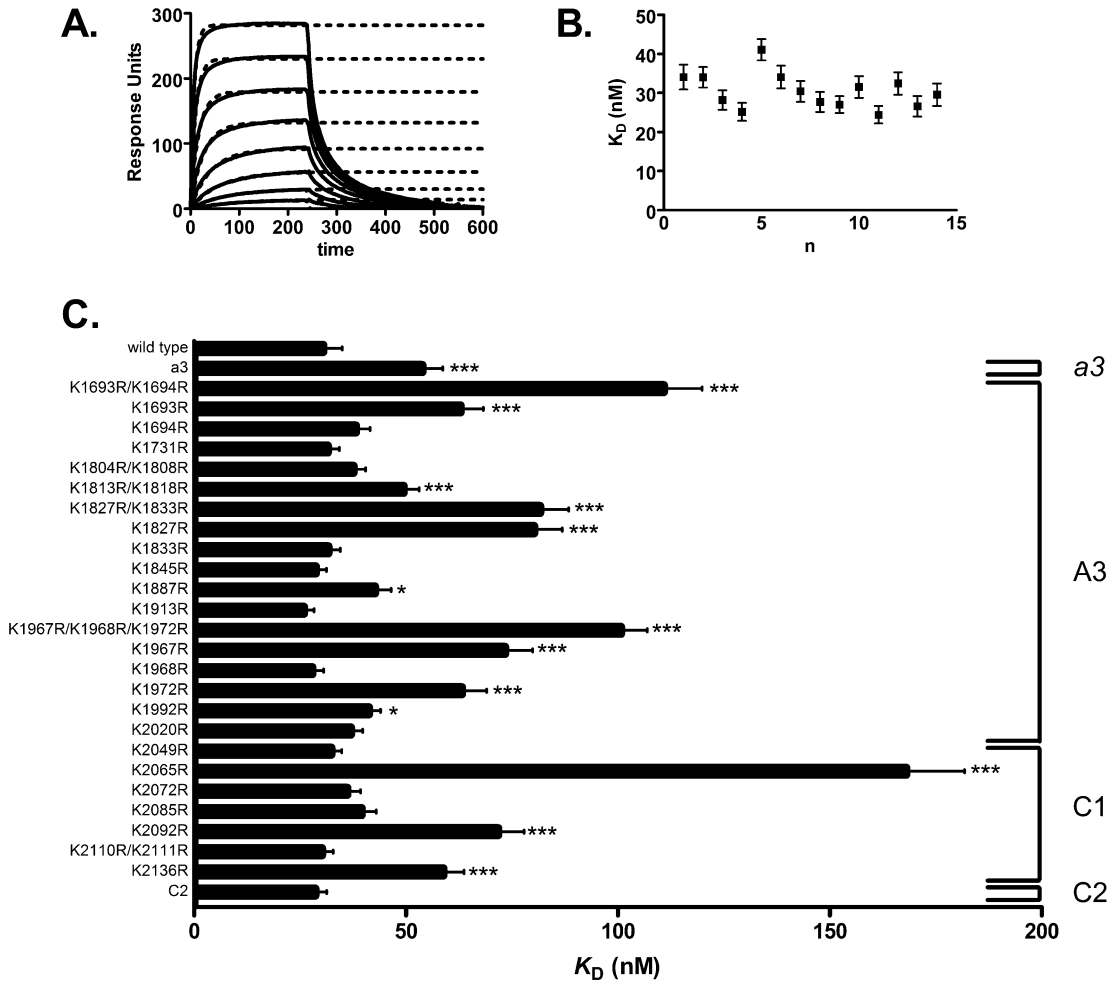

A.
Response Units
time
B.
K_D (nM)
n
C.
wild type
a3
K1693R/K1694R
K1693R
K1694R
K1731R
K1804R/K1808R
K1813R/K1818R
K1827R/K1833R
K1827R
K1833R
K1845R
K1887R
K1913R
K1967R/K1968R/K1972R
K1967R
K1968R
K1972R
K1992R
K2020R
K2049R
K2065R
K2072R
K2085R
K2092R
K2110R/K2111R
K2136R
C2
a3
A3
C1
C2
K_D (nM)

Figure 5.

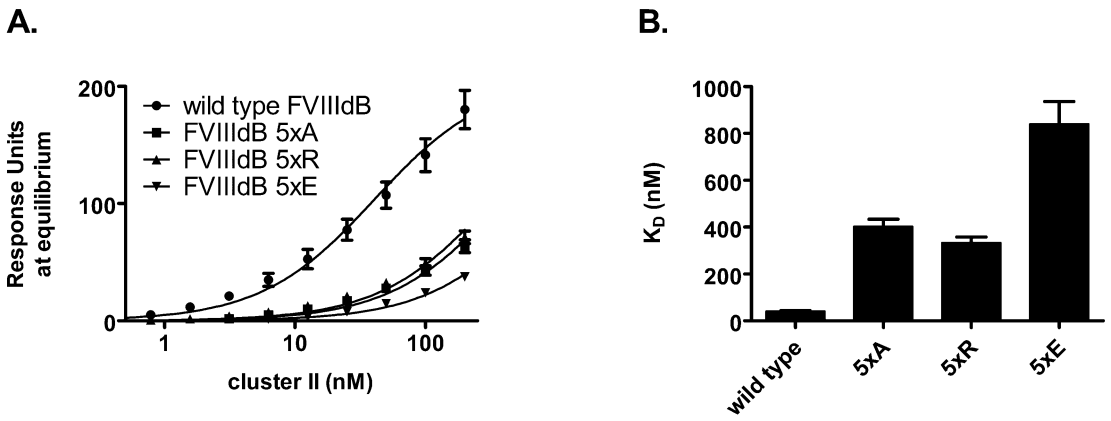
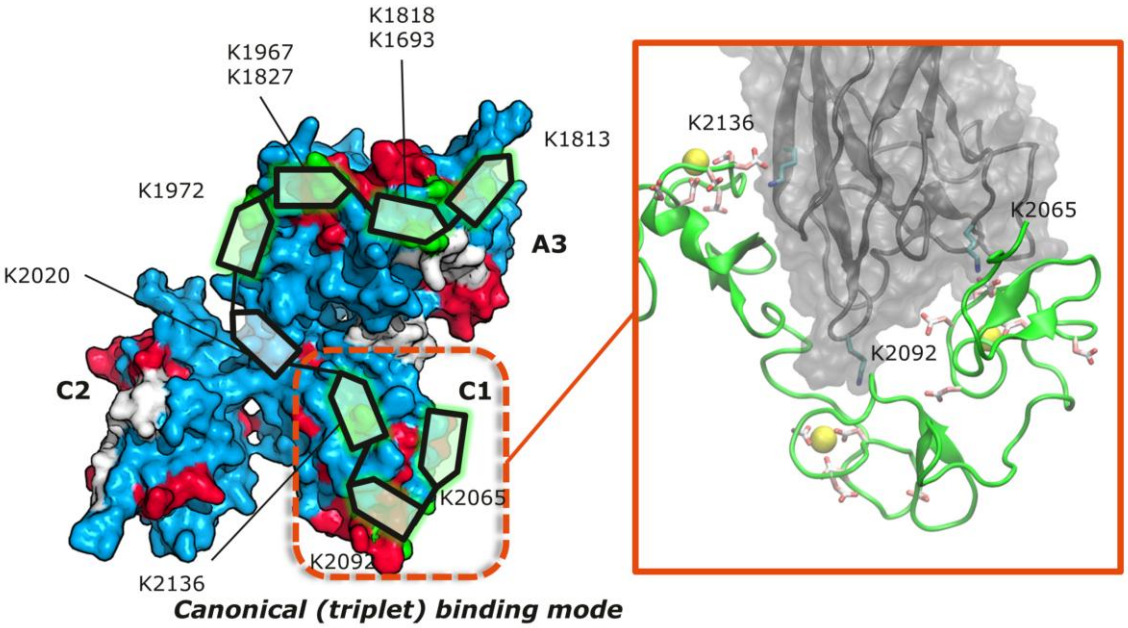
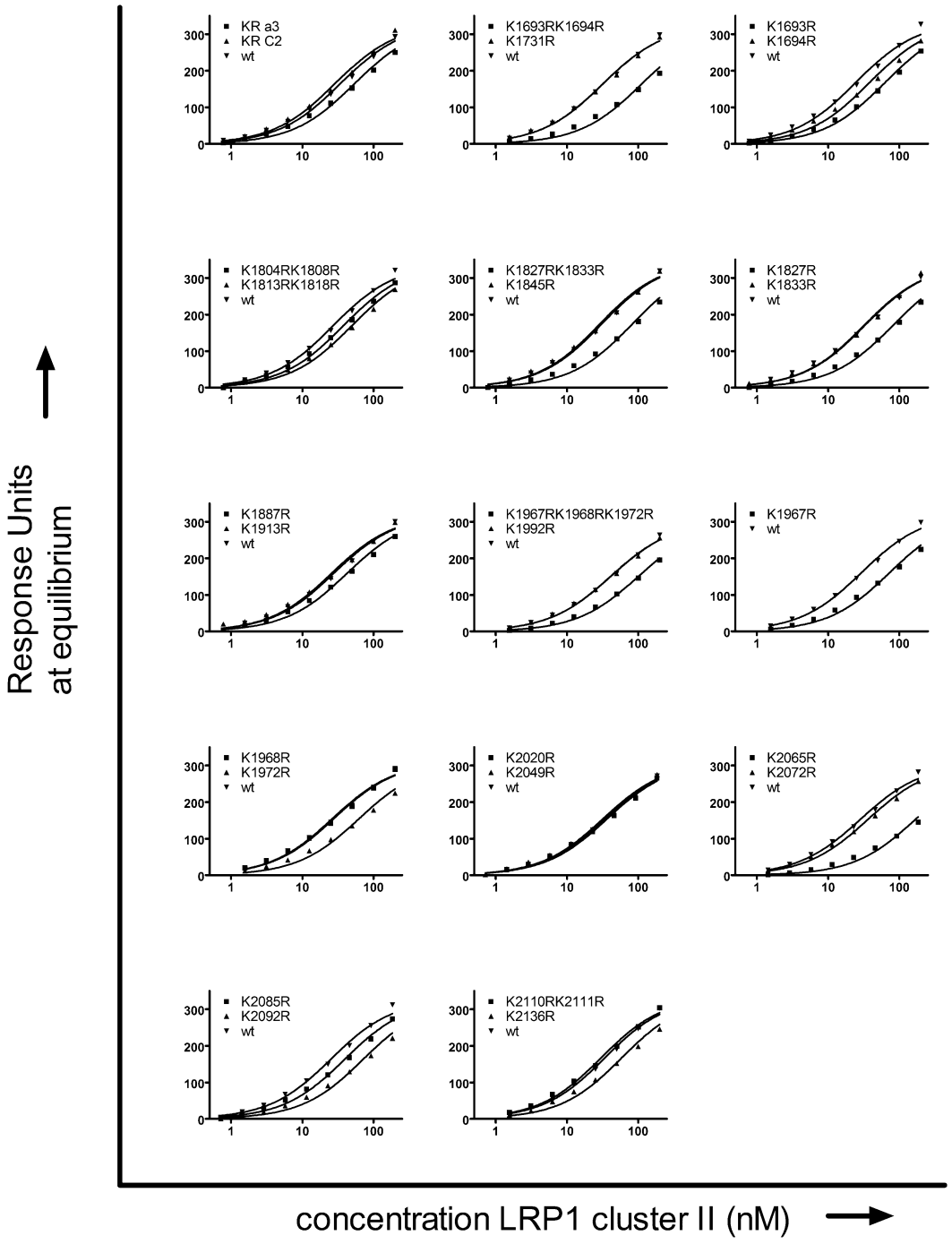

A.
B.
Response Units at equilibrium
200
100
0
1
10
100
cluster II (nM)
wild type FVIIIdB
FVIIIdB 5xA
FVIIIdB 5xR
FVIIIdB 5xE
K_D (nM)
1000
800
600
400
200
0
wild type
5xA
5xR
5xE

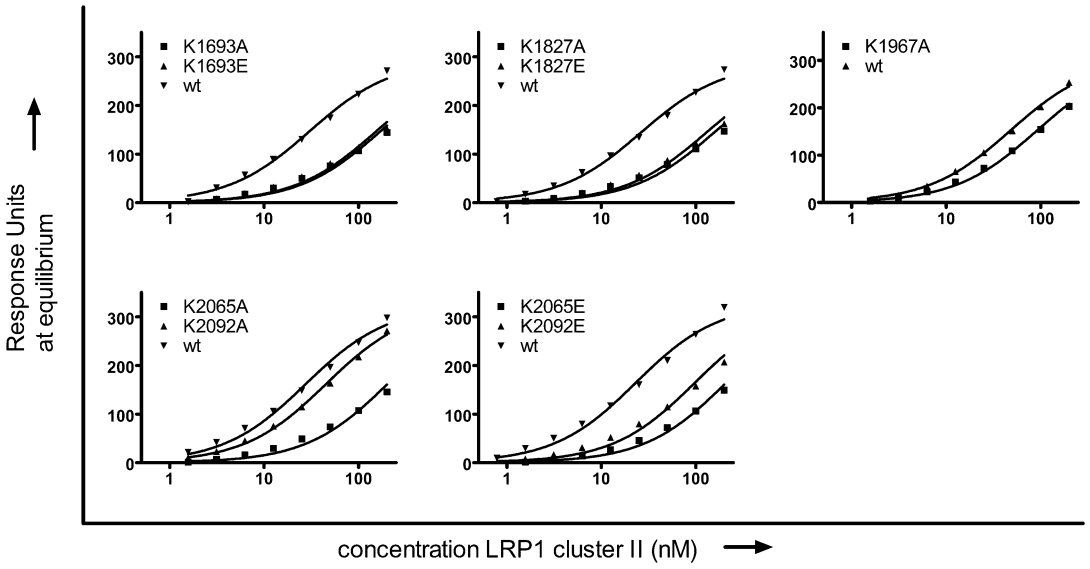
Figure 6.


K1818
K1693
K1967
K1827
K1813
K1972
A3
K2020
C2
C1
K2065
K2092
K2136
K2136
K2065
K2092
Canonical (triplet) binding mode

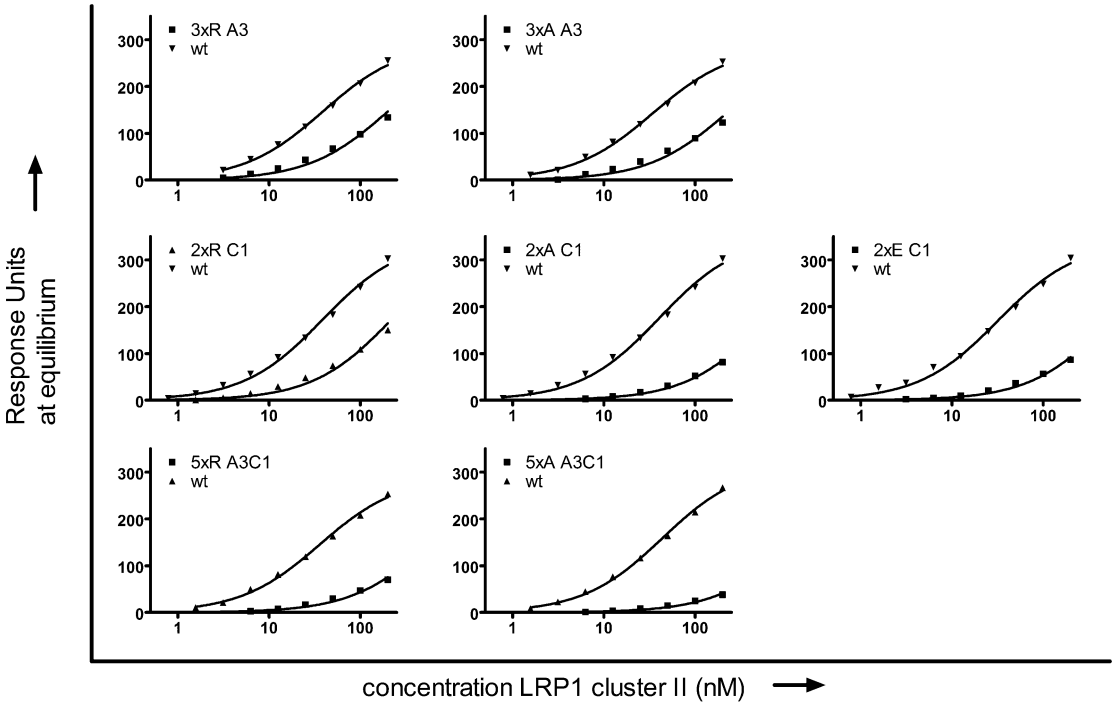


Supplemental Figure 1.

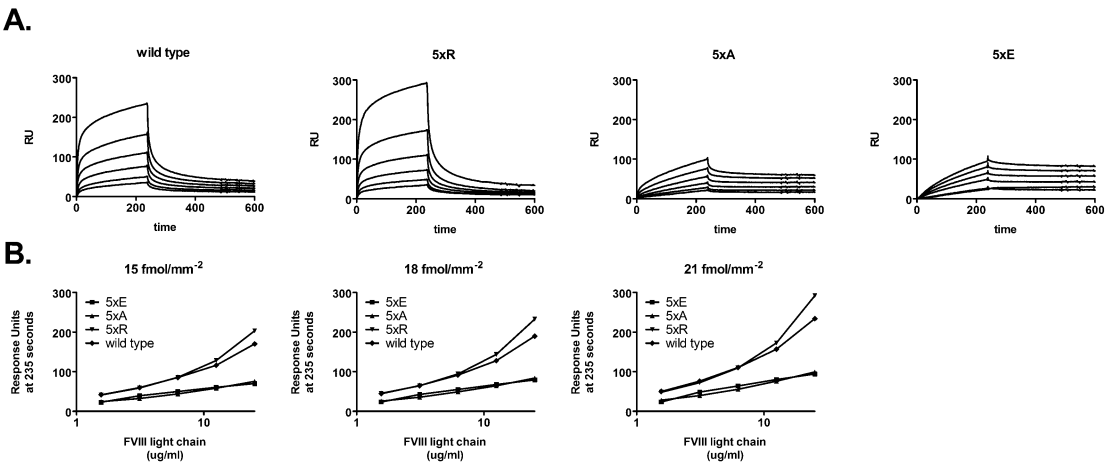
Supplemental Figure 2.



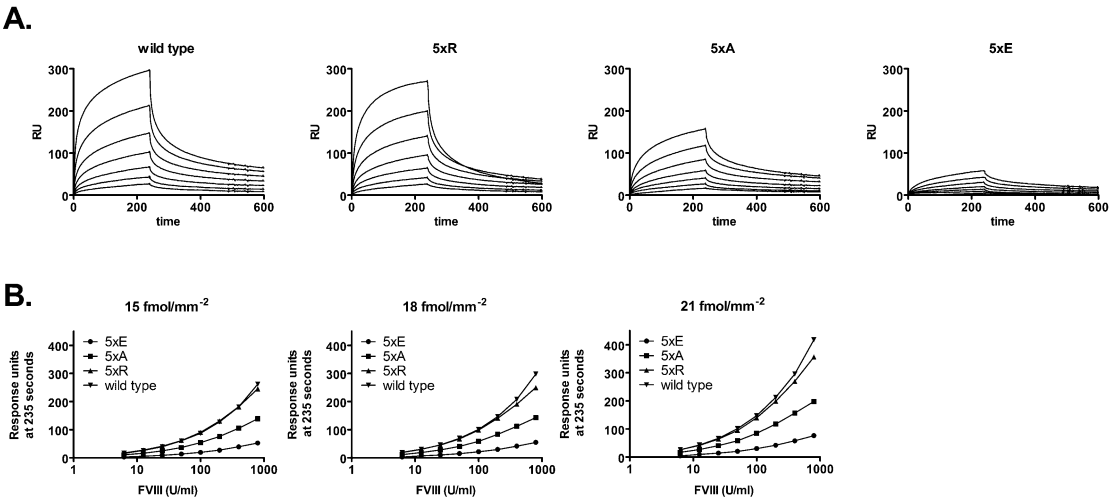
Supplemental Figure 3.



Supplemental Figure 4.



Supplemental Figure 5.



Supplemental methods

Structural Modeling

To construct a LRP model in consensus with the receptor ligand family binding motif however, we chose to create a homology model of LRP based on the LDLR-RAP complex as a template. This was done using the Rosetta Comparative Modeling protocol(51) by generating 30,000 decoys of the LRP1 CR56 double module. Calcium-ions, one per complementary repeat, were inserted based on the structural alignments and adjusted slightly by hand to remove atomistic overlap. The resulting calcium-loaded LRP double complementary repeat 56 (LRP1 CR56) model was positioned relative to the light chain of FVIII as to reconstruct the interaction motif of the LDLR-RAP complex in which CR5 interacts with K2065 and CR6 with K2092. The lysine residues of the FVIII light chain, K2065 and K2092 were overlaid with the NZ atoms of the two interacting lysine residues (K256 and K270) of RAP. To complete the triplet binding mode, a further CR5 module was positioned to interact with K2136 and connected to its adjacent CR56 module by the loopmodeling application in Rosetta, making 1,000 loop decoys. The specific sequence to be remodeled in each fragment was identified using the domain linker predictor DLP-SVM(52). The sequence for a CR56-CR56 chain was analyzed using their SVM-Joint consolidated predictor which identified region "CSH-SARTCPPNQFSCASGRCIPI" (residues 80-102 according to LRP alignment numbering) as a possible inter-domain linker. Since the natural chain break is after position 82 and both Cys 80 and Cys 87 are involved in forming intra-module disulfide bonds, we chose to limit the remodel region to residues 80-87. The resulting LRP1 CR565:FVIII light chain model in which the interaction pattern is CR565 against K2065-K2092-K2136 is truncated to remove the a3A3 and C2 domains of the FVIII light chain, leaving only the C1 domain (chain B residues 2020 to 2171). This system was solvated with ~10,000 TIP3P water molecules and ionized to salt concentration of 0.15. Following an energy minimization for 5,000 steps by a steepest decent method, the system was equilibrated using molecular dynamics simulations with constant number of atoms (N), constant pressure (P = 1 atm), and constant temperature (T = 310 K) for 5 ns, allowing it to settle.

5. Concluding remarks and perspectives

Molecular dynamics simulations have been performed to elaborate in atomistic detail intricate structural and functional relationships of constituting components of the two principal factor X-activating protein complexes in the blood coagulation cascade.

The coagulation cascade, though extensively studied, still has many un-explained subtleties and poorly understood concepts. Some of these are addressed using theoretical methods in this dissertation. In particular, the crystal structure of the un-inihibited FVII zymogen (ms.1) reveals that TF does not induce the rigidification of the activation (SP) domain because major parts are missing (and assumed to be too flexible). This flexibility was supported by the performed simulations of the homology model complex. Hence, the stabilization of the active site domain (including the formation of the S1 pocket and the oxyanion hole) by rigidification of especially the activation loops (1, 2, and 3) is most likely induced by the substrate (or inhibitor). Furthermore, we showed that TF could preserve the productive conformation of the activation domain in FVIIa (ms.2). This decisive result, in combination with the observations regarding the trypsin 170-loop-swapped variants, suggest for the first time an allosteric mechanism involving the E2 strand-following segment (residues 364{c215}-368{c219}) and putatively W364{c215} acting as a gating-residue for the active site cleft. The last study involving the extrinsic tenase complex looked at EGF1-EGF2 linker-deleted FVIIa variants and their ability to productively interact with tissue factor (ms.3). While it is long established that FVIIa binds TF in an extended conformation, is has not previously been shown how this extension related to the flexibility of said linker. Very recently was the inter-kringle domain linker investigated in prothrombin with respect to its role in the activation process.(Pozzi, Chen, Pelc, Shropshire, & Di Cera, 2014) The membrane-bound state of the FVIIIa molecule is important for appropriate co-factor function. Using a membrane-mimetic model known to facilitate the insertion of proteins, we studied the resulting conformations of the membrane-anchoring (discoidin) domains from FVIII (ms.4). Surprisingly, FVIII C1 and FVIII C2 did not converge to similar distribution of membrane-bound states emphasizing the subtleties and different roles of each domain in the binding process. Lastly, we

constructed an illustrative canonical binding mode between the LRP receptor and FVIII as ligand (ms.5).

The ideas and concepts explored in this dissertation have cousins in other areas of biochemistry due to the relative abundance of (serine) protease enzymatic systems. The fundamental functional construction of the coagulation system is a battery of serine proteases with auxiliary domains, regulators (cofactors and inhibitors), substrates, and receptors. This exact recipe is also adopted by the complement system (C1r, C1s and the C3 convertases e.g.) whereas the digestion system proteases (trypsin, chymotrypsin and elastase) are simpler molecules only consisting of the peptidase domain with activation peptides. Additionally, it is interesting to mention the subtilisin family of serine proteases. These share the exact same catalytic triad as serine proteases of the chymotrypsin family but have a completely different structural fold; a typical textbook example of a convergent result in evolution.

Throughout this dissertation, conventional (or classical) all-atomic MD simulations have been used as the *de facto* tool to investigate the mentioned structure-function relationships. It is, in this context, worth noting that it, for certain classes of problems, might be more appropriate to study the systems of interest at a different level of theory than chosen here. For instance, to do justice to e.g. the coordination chemistry of ions, a more detailed description at the quantum level is required. Multimerization of biomolecular components, on the other hand, is best studied using a more coarse grained description of the system, allowing simulation of larger time scales for even bigger systems. Yet another issue to philosophize about is sampling strategy. Basically, two schools of thought exist: 1) equilibrium simulations could be as long as possible or 2) parallel (much shorter) simulations could be executed. The caveat for the second procedure is obviously that this tactic greatly reduces the accessible time scales and, in consequence, slower processes would remain uncharted.

Pioneering work on the modelling of important haemostatic proteins and complexes has been performed by the group of Lee G. Pedersen. Using a combination of (sometimes incomplete) X-ray crystallographic structures, homology modelling, and experimentally

guided docking, Pedersen and co-workers have proposed models for the protein C zymogen(L Perera et al., 2000), the FIX zymogen(Lalith Perera, Darden, & Pedersen, 2001), the ternary prothrombinase:prothrombin FVa:FXa:PT complex(C J Lee, Wu, & Pedersen, 2011), and the binary tenase TF:FVIIa complex with TF containing the transmembrane part(Chang Jun Lee et al., 2012). However, a major challenge still stands to adequately sample the dynamical ensemble of such models in order to explain accumulating and increasingly complex experimental observations.

The studies presented in this thesis have added structural insight at the atomistic scale and a dynamical perspective on FVII activation and FVIII trafficking. Future challenges in the field include attempts to introduce the present findings into an ever more realistic physical and complex biological setting.

References

Adams, T. E., Hockin, M. F., Mann, K. G., & Everse, S. J. (2004). The crystal structure of activated protein C-inactivated bovine factor Va: Implications for cofactor function. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(24), 8918–23. doi:10.1073/pnas.0403072101

Allen, M. P., & Tildesley, D. J. (1989). *Computer simulation of liquids* (Reprint ed., p. 408). Oxford University Press.

Ash, E. L., Sudmeier, J. L., Day, R. M., Vincent, M., Torchilin, E. V, Haddad, K. C., … Bachovchin, W. W. (2000). Unusual 1H NMR chemical shifts support (His) C(epsilon) 1...O==C H-bond: proposal for reaction-driven ring flip mechanism in serine protease catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(19), 10371–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=27031&tool=pmcentrez&rendertype=abstract

Bajaj, S. P., Schmidt, A. E., Agah, S., Bajaj, M. S., & Padmanabhan, K. (2006). High resolution structures of p-aminobenzamidine- and benzamidine-VIIa/soluble tissue factor: unpredicted conformation of the 192-193 peptide bond and mapping of Ca2+, Mg2+, Na+, and Zn2+ sites in factor VIIa. *The Journal of Biological Chemistry*, *281*(34), 24873–88. doi:10.1074/jbc.M509971200

Banner, D. W., D'Arcy, A., Chène, C., Winkler, F. K., Guha, A., Konigsberg, W. H., … Kirchhofer, D. (1996). The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*, *380*, 41–46.

Bardelle, C., Furie, B., Furie, B. C., & Gilbert, G. E. (1993). Membrane binding kinetics of factor VIII indicate a complex binding process. *The Journal of Biological Chemistry*, *268*(12), 8815–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8473326

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., & MacKerell, A. D. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *Journal of Chemical Theory and Computation*, *8*, 3257–3273. doi:10.1021/ct300400x

Born, M., & Oppenheimer, R. (1927). Zur Quantentheorie der Molekeln. *Annalen Der Physik*, *4*(84), 457–484.

Bovenschen, N., Herz, J., Grimbergen, J. M., Lenting, P. J., Havekes, L. M., Mertens, K., & van Vlijmen, B. J. M. (2003). Elevated plasma factor VIII in a mouse model of low-density lipoprotein receptor–related protein deficiency. *Blood*, *101*(10), 3933–3939. doi:10.1182/blood-2002-07-2081

Brooks, B. R., Brooks III, C. L., MacKerell Jr, A. D., Nilsson, L., Petrella, R. J., Roux, B., … Karplus, M. (2009). CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry*, *30*(10), 1545–1614. doi:10.1002/jcc

Broze, G. J., & Majerus, P. W. (1980). Purification and properties of human coagulation factor VII. *Journal of Biological Chemistry*, *255*(4), 1242–1247. Retrieved from http://www.jbc.org/content/255/4/1242.short

Butenas, S., DiLorenzo, M. E., & Mann, K. G. (1997). Ultrasensitive fluorogenic substrates for serine proteases. *Thrombosis and Haemostasis*, *78*, 1193–1201. Retrieved from http://europepmc.org/abstract/MED/9364984

Car, R., & Parrinello, M. (1985). Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters*, *55*(22), 2471–2474. Retrieved from http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.55.2471

Davie, E. W., & Ratnoff, O. D. (1964). Waterfall Sequence for Intrinsic Blood Clotting. *Science (New York, N.Y.)*, *145*, 1310–1312.

Dennis, M. S., Eigenbrot, C., Skelton, N. J., Ultsch, M. H., Santell, L., Dwyer, M. a, … Lazarus, R. a. (2000). Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature*, *404*(6777), 465–70. doi:10.1038/35006574

Dijk, K. van, van der Bom, J. G., Lenting, P. J., de Groot, P. G., Mauser-Bunschoten, E. P., Roosendaal, G., … van den Berg, H. M. (2005). Factor VIII half-life and clinical phenotype of severe hemophilia A. *Haematologica*, *90*(4), 494–498. Retrieved from http://www.haematologica.org/content/90/4/494.short

Dolmer, K., Campos, A., & Gettins, P. G. W. (2013). Quantitative Dissection of the Binding Contributions of Ligand Lysines of the Receptor-associated Protein (RAP) to the Low Density Lipoprotein Receptor-related Protein (LRP1). *The Journal of Biological Chemistry*, *288*(33), 24081–90. doi:10.1074/jbc.M113.473728

Dolmer, K., & Gettins, P. G. W. (2006). Three complement-like repeats compose the complete alpha2-macroglobulin binding site in the second ligand binding cluster of the low density lipoprotein receptor-related protein. *The Journal of Biological Chemistry*, *281*(45), 34189–96. doi:10.1074/jbc.M604389200

Engelke, H. (2010). Coagulation protein FVIII binding to phospholipid membranes investigated by Fluorescence Correlation Spectroscopy.

Engelke, H., Lippok, S., Dorn, I., Netz, R. R., & Rädler, J. O. (2011). FVIII binding to PS membranes differs in the activated and non-activated form and can be shielded by annexin A5. *The Journal of Physical Chemistry. B*, *115*(44), 12963–70. doi:10.1021/jp2048579

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, *103*(19), 8577–8592. doi:10.1063/1.470117

Fisher, C., Beglova, N., & Blacklow, S. C. (2006). Structure of an LDLR-RAP complex reveals a general mode for ligand recognition by lipoprotein receptors. *Molecular Cell*, *22*(2), 277–83. doi:10.1016/j.molcel.2006.02.021

Frenkel, D., & Berend, S. (2001). *Understanding Molecular Simulation: From Algorithms to Applications (Computational Science)* (2nd revise., p. 664). Academic Press Inc.

Gandhi, P. S. (2010). Structural And Functional Studies On Thrombin Allostery. *Electronic Theses and Dissertations*, *Paper 848*.

Gandhi, P. S., Chen, Z., Mathews, F. S., & Di Cera, E. (2008). Structural identification of the pathway of long-range communication in an allosteric enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(6), 1832–7. doi:10.1073/pnas.0710894105

Gohara, D. W., & Di Cera, E. (2011). Allostery in trypsin-like proteases suggests new therapeutic strategies. *Trends in Biotechnology*, *29*(11), 577–85. doi:10.1016/j.tibtech.2011.06.001

Gomis-Rüth, F. X., Botelho, T. O., & Bode, W. (2012). A standard orientation for metallopeptidases. *Biochimica et Biophysica Acta*, *1824*(1), 157–63. doi:10.1016/j.bbapap.2011.04.014

Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chemical Reviews*, *102*(12), 4501–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12475199

Hilgartner, M. W., & Miller, D. R. (1975). The Effect of Cyanate on the Clotting Proteins and Platelet Function. *Experimental Biology and Medicine*, *149*(1), 5–9. doi:10.3181/00379727-149-38732

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics*, *14*, 33–38.

Huntington, J. A. (2008). How Na+ activates thrombin–a review of the functional and structural data. *Biological Chemistry*, *389*(8), 1025–1035. Retrieved from http://www.degruyter.com/view/j/bchm.2008.389.issue-8/bc.2008.113/bc.2008.113.xml

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935. doi:10.1063/1.445869

Kossiakoff, A. A., & Spencer, S. A. (1981). Direct determination of the protonation states of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine proteases: neutron structure of trypsin. *Biochemistry*, *20*(22), 6462–74. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7030393

Kumar, S., & Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *Journal of Molecular Biology*, *293*(5), 1241–55. doi:10.1006/jmbi.1999.3218

Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., … Bradley, P. (2011). ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology*, *487*(11), 545–74. doi:10.1016/B978-0-12-381270-4.00019-6

Lechtenberg, B. C., Murray-Rust, T. A., Johnson, D. J., Adams, T. E., Krishnaswamy, S., Camire, R. M., & Huntington, J. A. (2013). Crystal structure of the prothrombinase complex from the venom of Pseudonaja textilis. *Blood*, *122*(16), 2777–83. doi:10.1182/blood-2013-06-511733

Lee, C. J., Wu, S., Bartolotti, L. J., & Pedersen, L. G. (2012). Molecular dynamics simulations of the binary complex of human tissue factor (TF(1-242)) and factor VIIa (TF(1-242) /fVIIa) on a 4:1 POPC/POPS lipid bilayer. *Journal of Thrombosis and Haemostasis*, *10*, 2402–2405. doi:10.1111/j.1538-7836.2012.04920.x

Lee, C. J., Wu, S., & Pedersen, L. G. (2011). A proposed ternary complex model of prothrombinase with prothrombin: protein-protein docking and molecular dynamics

simulations. *Journal of Thrombosis and Haemostasis : JTH*, *9*(10), 2123–6. doi:10.1111/j.1538-7836.2011.04463.x

Lenting, P. J., VAN Schooten, C. J. M., & Denis, C. V. (2007). Clearance mechanisms of von Willebrand factor and factor VIII. *Journal of Thrombosis and Haemostasis*, *5*(7), 1353–60. doi:10.1111/j.1538-7836.2007.02572.x

Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, *61*(4), 704–21. doi:10.1002/prot.20660

MacFarlane, R. (1964). An enzyme cascade in the blood clotting mechanism, and its function as a biochemical amplifier. *Nature*, *202*, 498–499. Retrieved from http://www.nature.com/nature/journal/v202/n4931/abs/202498a0.html

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., … Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, *102*, 3586–3616.

Mandell, D. J., Coutsias, E. A., & Kortemme, T. (2009). Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature Methods*, *6*(8), 551–552. doi:10.1038/nmeth0809-551.Sub-angstrom

Marder, V. J., Aird, W. C., Bennett, J. S., Schulman, S., & White, G. C. (2012). *Hemostasis and Thrombosis: Basic Principles and Clinical Practice* (Sixth edit., p. 1592). LWW.

McCammon, J. A., & Harvey, S. C. (1988). *Dynamics of Proteins and Nucleic Acids* (New.). Cambridge University Press.

Meijer, A. B., Rohlena, J., van der Zwaan, C., van Zonneveld, A.-J., Boertjes, R. C., Lenting, P. J., & Mertens, K. (2007). Functional duplication of ligand-binding domains within low-density lipoprotein receptor-related protein for interaction with receptor associated protein, alpha2-macroglobulin, factor IXa and factor VIII. *Biochimica et Biophysica Acta*, *1774*(6), 714–22. doi:10.1016/j.bbapap.2007.04.003

Morrissey, J. H. (2005). Clotting Assay for Measuring Tissue Factor Activity. Retrieved from http://tf7.org/clot-assay.pdf

Mosnier, L. O., Zlokovic, B. V, & Griffin, J. H. (2007). The cytoprotective protein C pathway. *Blood*, *109*(8), 3161–3172. doi:10.1182/blood-2006-09-003004.

Mugherli, L., Burchak, O. N., Chatelain, F., & Balakirev, M. Y. (2006). Fluorogenic ester substrates to assess proteolytic activity. *Bioorganic & Medicinal Chemistry Letters*, *16*(17), 4488–91. doi:10.1016/j.bmcl.2006.06.037

Mulliken, R. S. (1955). Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *Journal of Chemical Physics*, *23*, 1833–1840. doi:http://dx.doi.org/10.1063/1.1740588

Ngo, J. C. K., Huang, M., Roth, D. A., Furie, B. C., & Furie, B. (2008). Crystal structure of human factor VIII: implications for the formation of the factor IXa-factor VIIIa complex. *Structure (London, England : 1993)*, *16*(4), 597–606. doi:10.1016/j.str.2008.03.001

Nibbering, E. T. J., Dreyer, J., Oliver, K., Bredenbeck, J., Hamm, P., & Elsaesser, T. (2014). *Vibrational dynamics of hydrogen bonds*. *Notes* (pp. 619–687).

Ohkubo, Y. Z., Morrissey, J. H., & Tajkhorshid, E. (2010). Dynamical view of membrane binding and complex formation of human factor VIIa and tissue factor. *Journal of Thrombosis and Haemostasis*, *8*, 1044–1053. doi:10.1111/j.1538-7836.2010.03826.x

Ohkubo, Y. Z., Pogorelov, T. V, Arcario, M. J., Christensen, G. A., & Tajkhorshid, E. (2012). Accelerating membrane insertion of peripheral proteins with a novel membrane mimetic model. *Biophysical Journal*, *102*(9), 2130–9. doi:10.1016/j.bpj.2012.03.015

Orcutt, S. J., & Krishnaswamy, S. (2004). Binding of substrate in two conformations to human prothrombinase drives consecutive cleavage at two sites in prothrombin. *The Journal of Biological Chemistry*, *279*(52), 54927–36. doi:10.1074/jbc.M410866200

Page, M. J., & Di Cera, E. (2010). Combinatorial enzyme design probes allostery and cooperativity in the trypsin fold. *Journal of Molecular Biology*, *399*(2), 306–319. doi:10.1016/j.jmb.2010.04.024

Patthy, L. (1985). Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. *Cell*, *41*(3), 657–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3891096

Perera, L., Darden, T. A., & Pedersen, L. G. (2001). Modeling human zymogen factor IX. *Thrombosis and Haemostasis*, (4), 596–603. Retrieved from http://europepmc.org/abstract/MED/11341491

Perera, L., Foley, C., Darden, T. a, Stafford, D., Mather, T., Esmon, C. T., & Pedersen, L. G. (2000). Modeling zymogen protein C. *Biophysical Journal*, *79*(6), 2925–43. doi:10.1016/S0006-3495(00)76530-1

Persson, E., Nielsen, L. S., & Olsen, O. H. (2001). Substitution of aspartic acid for methionine-306 in factor VIIa abolishes the allosteric linkage between the active site and the binding interface with tissue factor. *Biochemistry*, *40*, 3251–3256.

Persson, E., & Olsen, O. H. (2011). Allosteric activation of coagulation factor VIIa. *Frontiers in Bioscience*, *16*, 3156–3163.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., … Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*, 1781–1802. doi:10.1002/jcc.20289

Pike, A. C. W., Brzozowski, A. M., Roberts, S. M., Olsen, O. H., & Persson, E. (1999). Structure of human factor VIIa and its implications for the triggering of blood coagulation. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(16), 8925–30. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=17709&tool=pmcentrez&rendertype=abstract

Pozzi, N., Chen, Z., Pelc, L. a, Shropshire, D. B., & Di Cera, E. (2014). The linker connecting the two kringles plays a key role in prothrombin activation. *Proceedings of the National Academy of Sciences of the United States of America*, 1–6. doi:10.1073/pnas.1403779111

Rai, R., Kolesnikov, A., Sprengeler, P. A., Torkelson, S., Ton, T., Katz, B. A., … Young, W. B. (2006). Discovery of novel heterocyclic factor VIIa inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *16*(8), 2270–3. doi:10.1016/j.bmcl.2006.01.017

Rand, K. D., Andersen, M. D., Olsen, O. H., Jørgensen, T. J. D., Ostergaard, H., Jensen, O. N., … Persson, E. (2008). The origins of enhanced activity in factor VIIa analogs and the interplay between key allosteric sites revealed by hydrogen exchange mass spectrometry. *The Journal of Biological Chemistry*, *283*(19), 13378–87. doi:10.1074/jbc.M709716200

Rapaport, D. C. (2004). *The Art of Molecular Dynamics Simulation* (2nd ed., p. 564). Cambridge University Press.

Rauh, D., Klebe, G., & Stubbs, M. T. (2004). Understanding Protein–Ligand Interactions: The Price of Protein Flexibility. *Journal of Molecular Biology*, *335*(5), 1325–1341. doi:10.1016/j.jmb.2003.11.041

Ryckaert, J., Ciccotti, G., & Berendsen, H. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*, 327–341.

Saenko, E. L., Shima, M., & Sarafanov, A. G. (1999). Role of activation of the coagulation factor VIII in interaction with vWf, phospholipid, and functioning within the factor Xase complex. *Trends in Cardiovascular Medicine*, *9*(7), 185–92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10881749

Shannon, C. E. (1998). Communication In The Presence Of Noise. *Proceedings of the IEEE*, *86*(2), 447–457. doi:10.1109/JPROC.1998.659497

Sichler, K., Banner, D. W., D'Arcy, A., Hopfner, K.-P., Huber, R., Bode, W., … Brandstetter, H. (2002). Crystal Structures of Uninhibited Factor VIIa Link its Cofactor and Substrate-assisted Activation to Specific Interactions. *Journal of Molecular Biology*, *322*(3), 591–603. doi:10.1016/S0022-2836(02)00747-7

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(539), 539. doi:10.1038/msb.2011.75

Soejima, K., Mizuguchi, J., Yuguchi, M., Nakagaki, T., Higashi, S., & Iwanaga, S. (2001a). Factor VIIa modified in the 170 loop shows enhanced catalytic activity but does not change the zymogen-like property. *The Journal of Biological Chemistry*, *276*(20), 17229–35. doi:10.1074/jbc.M009206200

Soejima, K., Mizuguchi, J., Yuguchi, M., Nakagaki, T., Higashi, S., & Iwanaga, S. (2001b). Factor VIIa modified in the 170 loop shows enhanced catalytic activity but does not change the zymogen-like property. *The Journal of Biological Chemistry*, *276*(20), 17229–35. doi:10.1074/jbc.M009206200

Stierand, K., Maass, P. C., & Rarey, M. (2006). Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. *Bioinformatics (Oxford, England)*, *22*(14), 1710–6. doi:10.1093/bioinformatics/btl150

Stryer, L., Berg, J. M., & Tymoczko, J. L. (2002). *Biochemistry* (5th Revise., p. 1050). W.H.Freeman & Co Ltd.

The Uniprot Consortium. (2013). Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Research*, *41*, D43–7. doi:10.1093/nar/gks1068

Verlet, L. (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, *159*(1), 98–103.

Wang, J., Hao, P., & Li, Y. (2011). Exploration of conformational transition in the aryl-binding site of human FXa using molecular dynamics simulations. *Journal of Computational Chemistry*. doi:10.1007/s00894-011-1295-x

World Federation of Hemophilia. (2014). Retrieved from http://www.wfh.org/en/page.aspx?pid=642

Zakariassen, H., & Sørlie, M. (2007). Heat capacity changes in heme protein–ligand interactions. *Thermochimica Acta*, *464*(1-2), 24–28. doi:10.1016/j.tca.2007.07.021

N.B. References listed here are only for non-manuscript citations. Each manuscript is self-contained and as such includes a reference section.

Acknowledgements

Ole Hvilsted Olsen, principal scientist, Novo Nordisk A/S

for great inspirational drive and an always cheerful “can do” attitude.

Günther Herbert Peters, associate professor, DTU

for supervision and collaborative computational studies leading back to my early bachelor-degree years at DTU.

Johan Henrik Faber, head of department, Novo Nordisk A/S

for fruitful scientific discussions and motivational guidance.

Emadeedin Tajkhorshid & CSBMB group, Beckman Institute, UIUC

for harbouring my external stay at UIUC in the summer of 2013.

Yoshiaki Zenmei Ohkubo, research assistant, UIUC

for taking care of me during my stay in Urbana-Champaign, IL, (inlucding aiding me with the practicalities of said visit) and productive scientific discussions.

Maartje van den Biggelaar, research scientist, Sanquin Research

for intriguing scientific discussions at the quarterly Novo Nordisk-Sanquin meetings.

Dept. 416, haemophilia biochemistry, Novo Nordisk A/S

for coffee room discussions on scientific and non-scientific subjects alike as well as the entertaining (albeit rare) off-site events.